

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2005, 01:29:10 ; Search time 14083 Seconds
(without alignments)
17549.906 Million cell updates/sec

Title: US-10-768-511-5
Perfect score: 4348
Sequence: 1 atggcgccgcactcacgt.....gcgaattcccagagctgcg 4348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_scs.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_hg.*
15: gb_pi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4348	100.0	4348	6 AR490068	AR490068 Sequence
2	4348	100.0	4348	6 AX281543	AX281543 Sequence
3	673	15.5	723	6 AR490065	AR490065 Sequence
4	673	15.5	723	6 AX281540	AX281540 Sequence
5	520.8	12.0	3667	15 SPDCD21	X58824 Schizosacch
6	492.4	11.3	153751	2 AC116551	AC116551 Dictyoste
7	480.2	11.0	2733	6 AX489180	AX489180 Sequence
8	471.4	10.8	110000	14 CR954214_1	Continuation (2 of
9	470.8	10.8	5392	15 SCU14731	U14731 Saccharomyc
10	470.8	10.8	27166	15 SC9531X	Z49919 S.cerevisia
11	470.8	10.8	165536	15 SC9R3VI	Z71255 S.cerevisia
12	464.4	10.7	110000	15 CR380957_02	Continuation (3 of
13	458.6	10.5	110000	15 CR382132_07	Continuation (8 of
14	457.8	10.5	110000	15 AEO16817_06	Continuation (7 of
15	456	10.5	110000	15 CR382123_15	Continuation (16 o
16	444.2	10.2	3290	5 BC074670	BC074670 Xenopus t
17	436.8	10.0	3048	5 BC072870	BC072870 Xenopus l
18	435.2	10.0	2970	5 XLU46131	U46131 Xenopus lae

19	435.2	10.0	3051	5 XLU29178	U29178 Xenopus lae
20	430.8	9.9	3273	6 CQ981174	CQ981174 Sequence
21	430.8	9.9	3273	6 AX333086	AX333086 Sequence
22	430.8	9.9	3273	6 AX333087	AX333087 Sequence
23	430.8	9.9	3273	6 AX333494	AX333494 Sequence
24	430.8	9.9	3273	8 HSP1CDC21	X74794 H.sapiens p
25	430.8	9.9	3273	10 G28594	G28594 human STS S
26	430.4	9.9	3394	6 AX281814	AX281814 Sequence
27	430	9.9	2893	5 BC083031	BC083031 Xenopus l
28	429.2	9.9	2851	8 BC031061	BC031061 Homo sapi
29	427.6	9.8	2860	6 BD159700	BD159700 Primer fo
30	427.6	9.8	2860	6 AX882395	AX882395 Sequence
31	427.6	9.8	2860	8 AX022899	AX022899 Homo sapi
32	425.2	9.8	3087	5 XLU404049	U40409 Xenopus lae
33	413	9.5	110000	15 CR382137_13	Continuation (14 o
34	407.8	9.4	3042	5 BC065958	BC065958 Danio rer
35	407.6	9.4	2556	6 CQ717883	CQ717883 Sequence
36	395.8	9.1	3040	5 BC056514	BC056514 Danio rer
37	392.8	9.0	3221	9 MUSMCD21	D26089 Mus musculu
38	392.8	9.0	3283	9 AK220442	AK220442 Mus muscu
39	391.2	9.0	3360	9 BC013094	BC013094 Mus muscu
40	384.6	8.8	244538	14 AC129365	AC129365 Rattus no
41	381.6	8.8	2776	2 AK114961	AK114961 Ciona int
42	371.6	8.5	24140	14 AC119094	AC119094 Rattus no
43	329.6	7.6	2776	2 S80255	S80255 dpa-disc pr
44	329	7.6	2822	2 BT021329	BT021329 Drosophi
45	328.8	7.6	40751	15 SPCC24B10	AL157991 S.pombe c

ALIGNMENTS

RESULT 1	AR490068	Sequence 5 from patent US 6710229.	4348 bp	DNA	linear	PAT 15-MAY-2004
LOCUS	AR490068	Sequence 5 from patent US 6710229.	4348 bp	DNA	linear	PAT 15-MAY-2004
DEFINITION	AR490068	Sequence 5 from patent US 6710229.	4348 bp	DNA	linear	PAT 15-MAY-2004
ACCESSION	AR490068	Sequence 5 from patent US 6710229.	4348 bp	DNA	linear	PAT 15-MAY-2004
VERSION	AR490068.1	GI:47257191	4348 bp	DNA	linear	PAT 15-MAY-2004
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 4348)					
AUTHORS	da Costa e Silva,O., Bohnert,H.J., van Thielens,N., Chen,R. and Sarria-Willan,R.					
TITLE	Cell cycle stress-related proteins and methods of use in plants					
JOURNAL	Patent: US 6710229-A 5 23-MAR-2004;					
BASES	Plant Science GmbH; Ludwigshafen;					
WOX;						
FEATURES	Location/Qualifiers					
source	1..4348					
/organism="unknown"						
/mol_type="genomic DNA"						
ORIGIN						
Query Match	100.0%;	Score 4348;	DB 6;	Length 4348;		
Best Local Similarity	100.0%;	Pred No. 0;				
Matches 4348;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	ATGGCGCGCGCAGCTCACGTGAGGAATTCACCTCTCTGTTCTGCGACGTTCCATCTT	60			
Db	1	ATGGCGCGCGCAGCTCACGTGAGGAATTCACCTCTCTGTTCTGCGACGTTCCATCTT	60			
QY	61	TTTGGTTTTAGTTTGCAAACTCTGATCTGAGTTGAGAAAAGGCGGTTGTTGCT	120			
Db	61	TTTGGTTTTAGTTTGCAAACTCTGATCTGAGTTGAGAAAAGGCGGTTGTTGCT	120			
QY	121	TGAGGTGTTCTGTTGATTTGTCATCGAATAATGATGACATTCATTTGAGAGCGG	180			
Db	121	TGAGGTGTTCTGTTGATTTGTCATCGAATAATGATGACATTCATTTGAGAGCGG	180			
QY	181	TGTGTCGCCCATATCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACT	240			
Db	181	TGTGTCGCCCATATCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACT	240			

2401 CGGTGCACATCGGTGAACCTCAACCCCTTCAGATATAGACAAATGGTTCTCTGTTAAAGGAA 2460
Db |||||
2401 CGGTGCACATCGGTGAACCTCAACCCCTTCAGATATAGACAAATGGTTCTCTGTTAAAGGAA 2460
Qy |||||
2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAATTAAGGGGCCCTCTTCAAAATGTTTAG 2520
Db |||||
2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAATTAAGGGGCCCTCTTCAAAATGTTTAG 2520
Qy |||||
2521 TGTGTGGTCACTCGCCTCCGCTAGTTTACAGTTGTTTAAAGGGCGGTTGAGGAGCCAA 2580
Db |||||
2521 TGTGTGGTCACTCGCCTCCGCTAGTTTACAGTTGTTTAAAGGGCGGTTGAGGAGCCAA 2580
Qy |||||
2581 GGTGTGAAAAGCCAGATGTGCAGCAGCGAATGCTATGCTCTTATTCACAAATCGATGCA 2640
Db |||||
2581 GGTGTGAAAAGCCAGATGTGCAGCAGCGAATGCTATGCTCTTATTCACAAATCGATGCA 2640
Qy |||||
2641 CTTTGTGCAATTAAGCAGATAGTGCCTCTTCAAGAACTCCAGATGCCATTCTCTGAAGGAG 2700
Db |||||
2641 CTTTGTGCAATTAAGCAGATAGTGCCTCTTCAAGAACTCCAGATGCCATTCTCTGAAGGAG 2700
Qy |||||
2701 AGACTCCACACACAGTCAAGTGTGTTTATACCACTATGTTGATGCTGTGAAGCCTG 2760
Db |||||
2701 AGACTCCACACACAGTCAAGTGTGTTTATACCACTATGTTGATGCTGTGAAGCCTG 2760
Qy |||||
2761 GAGATCGTATTGAGGTAAACAGAGTTTCAAGGCCATGCGAGTTCGAGTTGGTCCGAATC 2820
Db |||||
2761 GAGATCGTATTGAGGTAAACAGAGTTTCAAGGCCATGCGAGTTCGAGTTGGTCCGAATC 2820
Qy |||||
2821 AACGAACATTACAGCATTTGATATAGACCTACATCGATTGCGTCAAGTCAAGATCTG 2880
Db |||||
2821 AACGAACATTACAGCATTTGATATAGACCTACATCGATTGCGTCAAGTCAAGATCTG 2880
Qy |||||
2881 ACAGGGTGCATGCAAACTGAAGATCTTATGAGATGATGAAGAGATGATATGATG 2940
Db |||||
2881 ACAGGGTGCATGCAAACTGAAGATCTTATGAGATGATGAAGAGATGATATGATG 2940
Qy |||||
2941 CTGGGTATCATGAAGTGTACTTTCAGAAAGCTGCTAATGAAGCAAGATTCAAAACTTA 3000
Db |||||
2941 CTGGGTATCATGAAGTGTACTTTCAGAAAGCTGCTAATGAAGCAAGATTCAAAACTTA 3000
Qy |||||
3001 AAGAGTGTCCAAGTCCCGGGCAATTTATGATAGACTTTCAAGGTGCGTGGCTCCAGCA 3060
Db |||||
3001 AAGAGTGTCCAAGTCCCGGGCAATTTATGATAGACTTTCAAGGTGCGTGGCTCCAGCA 3060
Qy |||||
3061 TTTGGAGCTTGAGATATTAAGAGGTCTTCTTGGCCAGCTCTTTGGTGGGAGGCTA 3120
Db |||||
3061 TTTGGAGCTTGAGATATTAAGAGGTCTTCTTGGCCAGCTCTTTGGTGGGAGGCTA 3120
Qy |||||
3121 AGAAAATTCATCTGAGCATCTTCCGAGGTGACATCAATGTTTACTTTTGGGGACC 3180
Db |||||
3121 AGAAAATTCATCTGAGCATCTTCCGAGGTGACATCAATGTTTACTTTTGGGGACC 3180
Qy |||||
3181 CTGGTACCAAGTAAATCTCAGTGTCTTCAAGTGTGCAAGATGCTCTCGTGGAACT 3240
Db |||||
3181 CTGGTACCAAGTAAATCTCAGTGTCTTCAAGTGTGCAAGATGCTCTCGTGGAACT 3240
Qy |||||
3241 ACATAGTGGCGAGGAAGTTCCGGCGTTGGGCTGACAGGTATGTAAACGAAGATCCAG 3300
Db |||||
3241 ACATAGTGGCGAGGAAGTTCCGGCGTTGGGCTGACAGGTATGTAAACGAAGATCCAG 3300
Qy |||||
3301 AAACCTGAGNAGCGGTATTTGGAGCGGAGCTTTGGTCTTCTAGTATCGTGGGATATGCT 3360
Db |||||
3301 AAACCTGAGNAGCGGTATTTGGAGCGGAGCTTTGGTCTTCTAGTATCGTGGGATATGCT 3360
Qy |||||
3361 GTATCGATGAGTTCGACAAAATGTCTGATATGCTGATAATGCCGAGCATGCTTCATGAGGTATGG 3420
Db |||||
3361 GTATCGATGAGTTCGACAAAATGTCTGATATGCTGATAATGCCGAGCATGCTTCATGAGGTATGG 3420
Qy |||||
3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCAATGCTCGTGAACGCTCGGACGT 3480
Db |||||
3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCAATGCTCGTGAACGCTCGGACGT 3480
Qy |||||
3481 CTGCTCTTGATGTGCAAAATCTAGTGGGTCCCGATCAATGCGCGCTTCTGTGATG 3540

Db |||||
3481 CTGCTCTTGATGTGCAAAATCTAGTGGGTCCCGATCAATGCGCGCTTCTGTGATG 3540
Qy |||||
3541 ATAAATCCAGTCTTCTCCAACTCTACTTTCTAGATTGATTTAAATTTACTTAATGCTCG 3600
Db |||||
3541 ATAAATCCAGTCTTCTCCAACTCTACTTTCTAGATTGATTTAAATTTACTTAATGCTCG 3600
Qy |||||
3601 ACAAACAGAGCAGCAAAACGATCGTCTGCGCCAGGCATCTCGTGGCTTTACACTATG 3660
Db |||||
3601 ACAAACAGAGCAGCAAAACGATCGTCTGCGCCAGGCATCTCGTGGCTTTACACTATG 3660
Qy |||||
3661 AAAACTATGAAGTTTCAAAGCAGCAGCCTTAGATCTACAAACACTTACCSCGTATATCA 3720
Db |||||
3661 AAAACTATGAAGTTTCAAAGCAGCAGCCTTAGATCTACAAACACTTACCSCGTATATCA 3720
Qy |||||
3721 CCTATGCTGTCAGCATGTATCATCTCATTAAGTGAAGAGTCTGCTGAAGTTGATTA 3780
Db |||||
3721 CCTATGCTGTCAGCATGTATCATCTCATTAAGTGAAGAGTCTGCTGAAGTTGATTA 3780
Qy |||||
3781 ATGGTATGTTGAGATGCGCCAAAAGGCACTTCTGGAAGCAGTAAAGGTTGATAA 3840
Db |||||
3781 ATGGTATGTTGAGATGCGCCAAAAGGCACTTCTGGAAGCAGTAAAGGTTGATAA 3840
Qy |||||
3841 CAGCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCCAATGA 3900
Db |||||
3841 CAGCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCCAATGA 3900
Qy |||||
3901 GATTTCTGAAAGTGTAGAGAAAGTTGATCAGCAGCAAGAGTGTGCGCCTTTTGAAGCTCG 3960
Db |||||
3901 GATTTCTGAAAGTGTAGAGAAAGTTGATCAGCAGCAAGAGTGTGCGCCTTTTGAAGCTCG 3960
Qy |||||
3961 CTTTGCAGCAATCTGCTACTCTGATCATGCAACAGAGTACGATAGACATGATCTTATCAGCA 4020
Db |||||
3961 CTTTGCAGCAATCTGCTACTCTGATCATGCAACAGAGTACGATAGACATGATCTTATCAGCA 4020
Qy |||||
4021 CTGGAGTCTCGGCGAGCAGAGGTATTCCTCGGCGCAACTTGTAGCTGCTCTCGCAGAGC 4080
Db |||||
4021 CTGGAGTCTCGGCGAGCAGAGGTATTCCTCGGCGCAACTTGTAGCTGCTCTCGCAGAGC 4080
Qy |||||
4081 TTATAGCAGATAAAATTTTCACTGGCAGCTCCTCTGGCTTGAAGACCAAGTCAAGTCTTTG 4140
Db |||||
4081 TTATAGCAGATAAAATTTTCACTGGCAGCTCCTCTGGCTTGAAGACCAAGTCAAGTCTTTG 4140
Qy |||||
4141 AGGATATCCGGAGCCAAAGCAGTGTGAGCTTTAGTTTTCAGAGATATTAATAATGCTCTGG 4200
Db |||||
4141 AGGATATCCGGAGCCAAAGCAGTGTGAGCTTTAGTTTTCAGAGATATTAATAATGCTCTGG 4200
Qy |||||
4201 GTAGCTTCCAGGAGAAAGCTTTCTTACTGTCAGTGTGACATAGTCAAGAGATTGAG 4260
Db |||||
4201 GTAGCTTCCAGGAGAAAGCTTTCTTACTGTCAGTGTGACATAGTCAAGAGATTGAG 4260
Qy |||||
4261 ACAGTTTCTTAACTGTTTGAATCCATGAGCTATAAATCTCTGAACGAAGGAAAACCTCCAG 4320
Db |||||
4261 ACAGTTTCTTAACTGTTTGAATCCATGAGCTATAAATCTCTGAACGAAGGAAAACCTCCAG 4320
Qy |||||
4321 TTTCCATGCGCAATTTCCAGAGCTCGC 4348
Db |||||
4321 TTTCCATGCGCAATTTCCAGAGCTCGC 4348

RESULT 2

AX281543

LOCUS

DEFINITION Sequence 5 from Patent WO0177354.

ACCESSION AX281543

VERSION AX281543.1

KEYWORDS GI:16608798

SOURCE

ORGANISM

Physcomitrella patens

Physcomitrella patens

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;

Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

REFERENCE 1

AUTHORS costa e Silva,O.D., Bohnert,H.J., van Thielens,N., Chen,R. and

DNA linear

PAT 02-NOV-2001

Sarria-Millan,R. Cell cycle stress-related proteins and methods of use in plants Patent: WO 0177354-A 5 18-OCT-2001; BASF Plant Science GmbH (DE) Location/Qualifiers 1..4348 /organism="Physcomitrella patens" /mol_type="unassigned DNA" /db_xref="taxon:3218"									
ORIGIN									
Query Match		100.0%;		Score 4348;		DB 6;		Length 4348;	
Best Local Similarity		100.0%;		Pred. No. 0;					
Matches 4348;		Conservative 0;		Mismatches 0;		Indels 0;		Gaps 0;	
QY	1	ATGCGCGCGGCACTCAGTGAAGAAATGCGACCTCCTTGTCTGCGACGGTTCATCTT	60						
Db	1	ATGCGCGCGGCACTCAGTGAAGAAATGCGACCTCCTTGTCTGCGACGGTTCATCTT	60						
QY	61	TTTGGTCTTTAGTTTGCATACTTCATCTGTGGAGTTGAGAAAGGGCGGTTCTGTGCT	120						
Db	61	TTTGGTCTTTAGTTTGCATACTTCATCTGTGGAGTTGAGAAAGGGCGGTTCTGTGCT	120						
QY	121	TGAGGTGTTCTTGTGATTTGCTCATGGAAATAATGATGCATCTGACATTTGGAGCGG	180						
Db	121	TGAGGTGTTCTTGTGATTTGCTCATGGAAATAATGATGCATCTGACATTTGGAGCGG	180						
QY	181	TGTCGTCGCCATATCCTTTGCGCAATCTGAAGAGTGTCTACGCCATTTGCCGAAGTAA	240						
Db	181	TGTCGTCGCCATATCCTTTGCGCAATCTGAAGAGTGTCTACGCCATTTGCCGAAGTAA	240						
QY	241	CACCGAGCTTCGACAAATCGAGCTCACCCTGCGCGGCGGAGGCGGTACGGCAGACC	300						
Db	241	CACCGAGCTTCGACAAATCGAGCTCACCCTGCGCGGCGGAGGCGGTACGGCAGACC	300						
QY	301	CTACATCTGCAGTTCGAAGGAGGAGAGAAACGGATTCGCTCGTAGGAGAGTGC	360						
Db	301	CTACATCTGCAGTTCGAAGGAGGAGAGAAACGGATTCGCTCGTAGGAGAGTGC	360						
QY	361	GATCTCGCAGTTTAGGCAATCTGTTATAGTTCCTTACGATCGGGGACTCCTGGAA	420						
Db	361	GATCTCGCAGTTTAGGCAATCTGTTATAGTTCCTTACGATCGGGGACTCCTGGAA	420						
QY	421	CTCCTGGAACTCCAGTGGCTACTCCGTTTACGCTACCCGAGTCGGTACACTATGGTA	480						
Db	421	CTCCTGGAACTCCAGTGGCTACTCCGTTTACGCTACCCGAGTCGGTACACTATGGTA	480						
QY	481	CCCCATCGTTCATCGTGCACGCCACAGTACAAACAGCGCAGTTCGTTCCACAGG	540						
Db	481	CCCCATCGTTCATCGTGCACGCCACAGTACAAACAGCGCAGTTCGTTCCACAGG	540						
QY	541	GGAAGCCTCTACATCGGAGACGTGATCTCAATCCAGAGAACCCGGGATCGATCTCCTT	600						
Db	541	GGAAGCCTCTACATCGGAGACGTGATCTCAATCCAGAGAACCCGGGATCGATCTCCTT	600						
QY	601	CAAGGGAACCTAGTCTGATGGGCTCCCTCTGAATCTGCTGAGCCAGATGACATTTGG	660						
Db	601	CAAGGGAACCTAGTCTGATGGGCTCCCTCTGAATCTGCTGAGCCAGATGACATTTGG	660						
QY	661	GTGGAGAAATATGCTTTATGTTGGGGGAGCAATGTTTAAATCCAGATGTCTAGGGCGA	720						
Db	661	GTGGAGAAATATGCTTTATGTTGGGGGAGCAATGTTTAAATCCAGATGTCTAGGGCGA	720						
QY	721	TTGTCGATTTTCCACAAATTAATCGTTCGAGTGTCTCATGATCTTAAATCCAGTACATCC	780						
Db	721	TTGTCGATTTTCCACAAATTAATCGTTCGAGTGTCTCATGATCTTAAATCCAGTACATCC	780						
QY	781	AGATCATAGAGAGACTGTGGAGCGTAGGAGGATCTCTAAATATGACATGTGAGACA	840						
Db	781	AGATCATAGAGAGACTGTGGAGCGTAGGAGGATCTCTAAATATGACATGTGAGACA	840						
QY	841	TTTATGACCATGATCTGATCTATACGCCAAATTTGTTTCGATPACCACTCCGACATCATCC	900						

Db	841	TTTATGACCATGATCTGATCTATACGCCAAATTTGTTTCGATATCCCATCTCGACATCATCC	900
QY	901	CCCTGTTGGACACTGAGTGTGAGAAAGTTGCTACTCTTTACTACCAAGCTTTGAGAGC	960
Db	901	CCCTGTTGGACACTGAGTGTGAGAAAGTTGCTACTCTTTACTACCAAGCTTTGAGAGC	960
QY	961	ATATTGAGGCGAGACCTTTTCAATCTCAAAGCATCGGTGCACATCGTGAACCTCAACCCCTT	1020
Db	961	ATATTGAGGCGAGACCTTTTCAATCTCAAAGCATCGGTGCACATCGTGAACCTCAACCCCTT	1020
QY	1021	CAGATATAGACAAATTTGTTTCTGTTAAAGAAATGTTTATCCGGTGCAGTGTCTATCATAC	1080
Db	1021	CAGATATAGACAAATTTGTTTCTGTTAAAGAAATGTTTATCCGGTGCAGTGTCTATCATAC	1080
QY	1081	CTGAAATTAAGGGGGCTTCTTCAAAATGTTTAGTGTGTGGTCACTCGGCTCGCTAGTTA	1140
Db	1081	CTGAAATTAAGGGGGCTTCTTCAAAATGTTTAGTGTGTGGTCACTCGGCTCGCTAGTTA	1140
QY	1141	CAGTTGTTTAAAGGGCGGTTGAGGAGCCAAACAAGGTGTGAAAGCCAGAAATGTGCAGCAC	1200
Db	1141	CAGTTGTTTAAAGGGCGGTTGAGGAGCCAAACAAGGTGTGAAAGCCAGAAATGTGCAGCAC	1200
QY	1201	GGAAATGCTATCTCTTTATTCACAAATCGATGCACCTTTTGCAAAATAAGCAGATAGTCGTC	1260
Db	1201	GGAAATGCTATCTCTTTATTCACAAATCGATGCACCTTTTGCAAAATAAGCAGATAGTCGTC	1260
QY	1261	TTCAAGAAACTTCCAGATGCGCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT	1320
Db	1261	TTCAAGAAACTTCCAGATGCGCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT	1320
QY	1321	TATACACACTATCGTTGATGCTGTGAAGCTGTGAGATCGTATTGAGGTAAACAGGAGTTT	1380
Db	1321	TATACACACTATCGTTGATGCTGTGAAGCTGTGAGATCGTATTGAGGTAAACAGGAGTTT	1380
QY	1381	TCAAGGCGCATGCGAGTTCGAGTTCGTCATGCGCGCGCGCACTCAGCTGAGGAATTCACCC	1440
Db	1381	TCAAGGCGCATGCGAGTTCGAGTTCGTCATGCGCGCGCGCACTCAGCTGAGGAATTCACCC	1440
QY	1441	TCCTTGTTCTCGGACGGTTCCATTCTTTTGGTTTTAGTTTGGCAAACTTTGATCGTGA	1500
Db	1441	TCCTTGTTCTCGGACGGTTCCATTCTTTTGGTTTTAGTTTGGCAAACTTTGATCGTGA	1500
QY	1501	GTTTCAGAAAAGGGGGTTCGTTCTTGTGAGTGTCTTGTGATTTGTTGTTGTTGTTGTTG	1560
Db	1501	GTTTCAGAAAAGGGGGTTCGTTCTTGTGAGTGTCTTGTGATTTGTTGTTGTTGTTGTTG	1560
QY	1561	ATAATGATGCACTTGACATTCGAGCCGTCGTCCCATATCTCTTCGCAATCTGAAGGAG	1620
Db	1561	ATAATGATGCACTTGACATTCGAGCCGTCGTCCCATATCTCTTCGCAATCTGAAGGAG	1620
QY	1621	TGTCTACGCCATTCGCGCAAGTAAATCAACCGAGTTTCGAAATTCGACGCTCACCCGTGG	1680
Db	1621	TGTCTACGCCATTCGCGCAAGTAAATCAACCGAGTTTCGAAATTCGACGCTCACCCGTGG	1680
QY	1681	CCGGCGGAGGGCGGTACGGCAGACCCCTCATCTGCAAGTTCGAAAGGAGGGAGAGAAA	1740
Db	1681	CCGGCGGAGGGCGGTACGGCAGACCCCTCATCTGCAAGTTCGAAAGGAGGGAGAGAAA	1740
QY	1741	CGGATTCGCTCGTGTAGGAGAGTTCGATCTCGCAGTTTAGGCAATTTCTGTTTATAGTT	1800
Db	1741	CGGATTCGCTCGTGTAGGAGAGTTCGATCTCGCAGTTTAGGCAATTTCTGTTTATAGTT	1800
QY	1801	CCCCTTACGATGCGGGGACTCCTGGAACTCCCTGAACTCCAGTGGCTACTCCGGTTTACG	1860
Db	1801	CCCCTTACGATGCGGGGACTCCTGGAACTCCCTGAACTCCAGTGGCTACTCCGGTTTACG	1860
QY	1861	CTACCCGATCGGTACACCTATGGTACCCCATCGTTCATTCGTCGCGCAGCCACAGTACA	1920
Db	1861	CTACCCGATCGGTACACCTATGGTACCCCATCGTTCATTCGTCGCGCAGCCACAGTACA	1920
QY	1921	AACAGCGCAGTGTGAGTGTGTTTCCAGGGGAGCCCTCTACATCGGAGAGCTCGATCTCAAT	1980
Db	1921	AACAGCGCAGTGTGAGTGTGTTTCCAGGGGAGCCCTCTACATCGGAGAGCTCGATCTCAAT	1980

QY	1981	CCAGAGAAACCGGGCATCGATCTCTTCAAGGGAACCTAGTGTCTGATGGGCGTCCCTCTG	2040	QY	3061	TTTGGAGCTTGAAGATATTAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAGGCTA	3120
DB	1981	CCAGAGAAACCGGGCATCGATCTCTTCAAGGGAACCTAGTGTCTGATGGGCGTCCCTCTG	2040	DB	3061	TTTGGAGCTTGAAGATATTAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAGGCTA	3120
QY	2041	AATCTGTGAGCCAGATGACATCTTGGGTGGGAATATGCTTATGTTGGGGGAGCAATG	2100	QY	3121	AGAAAAATCCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTATCTTGTGTGGGACC	3180
DB	2041	AATCTGTGAGCCAGATGACATCTTGGGTGGGAATATGCTTATGTTGGGGGAGCAATG	2100	DB	3121	AGAAAAATCCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTATCTTGTGTGGGACC	3180
QY	2101	TTAAACATTCAGATGTCTTGGGCGATTCGTCGATTTCTCCACAATATATCGTTCGAGTG	2160	QY	3181	CTGTGACACAGTAATCTCAGCTGTCTCAGTATGTGCAACAAGATAGCTCTCTGTGGAATCT	3240
DB	2101	TTAAACATTCAGATGTCTTGGGCGATTCGTCGATTTCTCCACAATATATCGTTCGAGTG	2160	DB	3181	CTGTGACACAGTAATCTCAGCTGTCTCAGTATGTGCAACAAGATAGCTCTCTGTGGAATCT	3240
QY	2161	CTCATGATCTTAATCCAGTACATCCAGATCATAGAGAGACTGTGGAGCGTGGAGGG	2220	QY	3241	ACACTAGTGGGCGAGGAATTTGGGCGTGGGCTGACAGCGTATGTAAAGAGATCCAG	3300
DB	2161	CTCATGATCTTAATCCAGTACATCCAGATCATAGAGAGACTGTGGAGCGTGGAGGG	2220	DB	3241	ACACTAGTGGGCGAGGAATTTGGGCGTGGGCTGACAGCGTATGTAAAGAGATCCAG	3300
QY	2221	ATACCTTAATATCGACATGTGAGACATTTATGACATGATCTCTATATACGCAAAA	2280	QY	3301	AAATCTCGAGAGACGGTATTGGAGCGGAGCTTTGGTCTTAGTGATCGTGGGATATGCT	3360
DB	2221	ATACCTTAATATCGACATGTGAGACATTTATGACATGATCTCTATATACGCAAAA	2280	DB	3301	AAATCTCGAGAGACGGTATTGGAGCGGAGCTTTGGTCTTAGTGATCGTGGGATATGCT	3360
QY	2281	TTGTTCGATACCACTCGACATCATCCCTCTGTGGACACTGAGTGTGAGGAGTTGCTA	2340	QY	3361	GTATCGATGAGTTTCGACAAAATGTCTGATTAATGCCGAGCATGCTTCATGAGGTAATGG	3420
DB	2281	TTGTTCGATACCACTCGACATCATCCCTCTGTGGACACTGAGTGTGAGGAGTTGCTA	2340	DB	3361	GTATCGATGAGTTTCGACAAAATGTCTGATTAATGCCGAGCATGCTTCATGAGGTAATGG	3420
QY	2341	CCTCTTTTACTACCAAGCTTTGAGAGCATATTTAGGCCAGACCTTTCAATCTCAAGCAT	2400	QY	3421	AGCAACAAACGGTATCTGTAGCCAAAAGGGGGTATCATTTGCCCTCGCTGAACGCTCGACCT	3480
DB	2341	CCTCTTTTACTACCAAGCTTTGAGAGCATATTTAGGCCAGACCTTTCAATCTCAAGCAT	2400	DB	3421	AGCAACAAACGGTATCTGTAGCCAAAAGGGGGTATCATTTGCCCTCGCTGAACGCTCGACCT	3480
QY	2401	CGGTGACATGCGTGAATCAACCTTTCAGATATAGACAAATTTGTTCTGTGTAAGGAA	2460	QY	3481	CTGTCTTGTGATGCGAAATCTAGTGGTCCCGATACAAATGCGGCGCTTCTGTGATATG	3540
BB	2401	CGGTGACATGCGTGAATCAACCTTTCAGATATAGACAAATTTGTTCTGTGTAAGGAA	2460	DB	3481	CTGTCTTGTGATGCGAAATCTAGTGGTCCCGATACAAATGCGGCGCTTCTGTGATATG	3540
QY	2461	TGGTTATCCGGTGAGTCTATCATACCTGAAATTAAGGGGCGCTTCTCAATGTTTAG	2520	QY	3541	ATAACATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTAATGCTCG	3600
DB	2461	TGGTTATCCGGTGAGTCTATCATACCTGAAATTAAGGGGCGCTTCTCAATGTTTAG	2520	DB	3541	ATAACATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTAATGCTCG	3600
QY	2521	TGTGTGTCACCTCGCTCGCTAGTTTAAAGGGGCGGTTGAGGAGCCAAACAA	2580	QY	3601	ACAAACACAGACGAGCAAAACGATCGTCTGCCAGGATCTCTGTGGCTTTACATATG	3660
DB	2521	TGTGTGTCACCTCGCTCGCTAGTTTAAAGGGGCGGTTGAGGAGCCAAACAA	2580	DB	3601	ACAAACACAGACGAGCAAAACGATCGTCTGCCAGGATCTCTGTGGCTTTACATATG	3660
QY	2581	GGTGTGAAAGCCAGAAATGTCAGCACGGAATGCTATGCTCTTATTCACAAATCGATGCA	2640	QY	3661	AAAACTATGAGTTTCAAAGCAGGACGCTTATAGATCTACAAACACTTCCGGGTATATCA	3720
DB	2581	GGTGTGAAAGCCAGAAATGTCAGCACGGAATGCTATGCTCTTATTCACAAATCGATGCA	2640	DB	3661	AAAACTATGAGTTTCAAAGCAGGACGCTTATAGATCTACAAACACTTCCGGGTATATCA	3720
QY	2641	CTTTTGCATTAAGCAGATAGTGCCTTTCAAGAACTCCAGATGCCATTTCTGGAAGGAG	2700	QY	3721	CTATGCTCGTCAGCATGTACATCTCAATTAAGTGTGATGAGCTGCTGAAGATTTGATTA	3780
DB	2641	CTTTTGCATTAAGCAGATAGTGCCTTTCAAGAACTCCAGATGCCATTTCTGGAAGGAG	2700	DB	3721	CTATGCTCGTCAGCATGTACATCTCAATTAAGTGTGATGAGCTGCTGAAGATTTGATTA	3780
QY	2701	AGACTCCACACAGTCAGATGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG	2760	QY	3781	ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCTCGAAGCAGTAAAGAGGTGATAA	3840
DB	2701	AGACTCCACACAGTCAGATGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG	2760	DB	3781	ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCTCGAAGCAGTAAAGAGGTGATAA	3840
QY	2761	GAGATCGTATTTGAGGTAAACAGGAGTTTCAAGGCCATGCGAGTTCGAGTTGGTCCGAATC	2820	QY	3841	GAGCCACACTCGGCAACTCGAAAGTATGATTCGATCAGTGAAGCCCTAGCTCGAATGA	3900
DB	2761	GAGATCGTATTTGAGGTAAACAGGAGTTTCAAGGCCATGCGAGTTCGAGTTGGTCCGAATC	2820	DB	3841	GAGCCACACTCGGCAACTCGAAAGTATGATTCGATCAGTGAAGCCCTAGCTCGAATGA	3900
QY	2821	AACGAACATTTACGAGCATTTGATAAGACCTTACATCGATTTGCGTGCACGTCAGAAAGTCTG	2880	QY	3901	GATTTTCTCAAGTGTGAGAGAAAGTTGATGTCAGCAGAAAGCTGTGCGCTTTTAGACGTG	3960
DB	2821	AACGAACATTTACGAGCATTTGATAAGACCTTACATCGATTTGCGTGCACGTCAGAAAGTCTG	2880	DB	3901	GATTTTCTCAAGTGTGAGAGAAAGTTGATGTCAGCAGAAAGCTGTGCGCTTTTAGACGTG	3960
QY	2881	ACAGGGTTCGACTGCAAACTGGAAGTCTTATGAGATGGAATAAGGAGATGATATGTATG	2940	QY	3961	CTTTGACAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA	4020
DB	2881	ACAGGGTTCGACTGCAAACTGGAAGTCTTATGAGATGGAATAAGGAGATGATATGTATG	2940	DB	3961	CTTTGACAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA	4020
QY	2941	CTGGGTATCATGAAAGTATCTTCAAGACTGCTAATGAAGCAAAAGATTCAAAAACTTA	3000	QY	4021	CTGGAGTGTGCGCCAGCGAGCTTATTCGTGCGGCCAACTTGTAGTGTCTCTCGGAGGC	4080
DB	2941	CTGGGTATCATGAAAGTATCTTCAAGACTGCTAATGAAGCAAAAGATTCAAAAACTTA	3000	DB	4021	CTGGAGTGTGCGCCAGCGAGCTTATTCGTGCGGCCAACTTGTAGTGTCTCTCGGAGGC	4080
QY	3001	AAGAGCTGTCCAAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTGCTGCTCCAGCA	3060	QY	4081	TTATAGCAGTAAAAATTTTCACTGCGCAGCTCTCTGCTGTTGAAGACAGCTCAGCTTCTTG	4140
DB	3001	AAGAGCTGTCCAAAGCTCCCGGGCATTTATGATAGACTTTTCAAGGTGCTGCTCCAGCA	3060	DB	4081	TTATAGCAGTAAAAATTTTCACTGCGCAGCTCTCTGCTGTTGAAGACAGCTCAGCTTCTTG	4140
				QY	4141	AGGATATCCGGAGCCAAAGCAGTGTGGAAGCTTAGTTTGCAGGATATTAAAAATGCTCTGG	4200

```
Db 4141 AGGATATCCGGAGCCAAAGCAGTGGAGCTTAGTTTCAGGATATTAATAATGCTCTGG 4200
QY 4201 GTAGCTCCAGGAGAGAGGCTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGG 4260
Db 4201 GTAGCTCCAGGAGAGAGGCTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGG 4260
QY 4261 ACAGTTTCTAACTGTTCCGAATCCATGAGCTATAACTCTGAAAGGAGGAAACCTCCAG 4320
Db 4261 ACAGTTTCTAACTGTTCCGAATCCATGAGCTATAACTCTGAAAGGAGGAAACCTCCAG 4320
QY 4321 TTTCCATGCGCAATPTCCAGAGCTCGC 4348
Db 4321 TTTCCATGCGCAATPTCCAGAGCTCGC 4348

RESULT 3
LOCUS AR490065 723 bp DNA linear PAT 15-MAY-2004
DEFINITION Sequence 2 from patent US 6710229.
ACCESSION AR490065
VERSION AR490065.1 GI:47257188
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 723)
AUTHORS da Costa e Silva,O., Bohnert,H.J., van Thielén,N., Chen,R. and Sarria-Millan,R.
TITLE Cell cycle stress-related proteins and methods of use in plants
JOURNAL Patent: US 6710229-A 2 23-MAR-2004; BASF Plant Science GmbH; Ludwigshafen; WOX;
FEATURES source
location/Qualifiers
1..723
/mol_type="unknown"
/organism="genomic DNA"
ORIGIN
Query Match 15.5%; Score 673; DB 6; Length 723;
Best Local Similarity 97.1%; Pred. No. 2.5e-189;
Matches 696; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 2753 GAAGCTGGAGATCGTATTGAGTTAACAGGAGTTTTCAGGCCATGGCAGTTCGAGTTGG 2812
Db 3 GCACAGGAGATCGTATTGAGTTAACAGGAGTTTTCAGGCCATGGCAGTTCGAGTTGG 62
QY 2813 TCCGAATCAACGAACATTACGAGCATTTGTATAAGACCTACATCGATTGCGTGACGTCAA 2872
Db 63 TCCGATCAACGAACATTACGAGCATTTGTATAAGACCTACATCGATTGCGTGACGTCAA 122
QY 2873 GAAGTCTGACAGGGTTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGATGA 2932
Db 123 GAAGTCTGACAGGGTTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGATGA 182
QY 2933 TATGTATCTGGGTATCATGAAGTGATACCTTCAAGACTGCTATGAAGCAAGATTCA 2992
Db 183 TATGTATCTGGGTATCATGAAGTGATACCTTCAAGACTGCTATGAAGCAAGATTCA 242
QY 2993 AAAAATTAAAGAGCTGCTCAAGCTCCCGGCACTTATGATAGACTTTTCAAGTTCGCTGGC 3052
Db 243 AAAAATTAAAGAGCTGCTCAAGCTCCCGGCACTTATGATAGACTTTTCAAGTTCGCTGGC 302
QY 3053 TCCAGCAATTTGGGAGCTTGAAGATATTAAGAGGCTCTCTTTGCCAGCTCTTTGGTGG 3112
Db 303 TCCAGCAATTTGGGAGCTTGAAGATATTAAGAGGCTCTCTTTGCCAGCTCTTTGGTGG 362
QY 3113 GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGTGACATCAATGTTTACTTGT 3172
Db 363 GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGTGACATCAATGTTTACTTGT 422
QY 3173 TGGGACCTCGTACAGTAATCTCAGCTGCTTCAGTATGTGCAAGATAGCTCTCTCG 3232
```

```
Db 423 TGGGACCTCGTACAGTAATCTCAGCTGCTTTCAGTATGTGCAAGATAGCTCTCTCG 482
QY 3233 TGGATCTACACTAGTGGCGAGGAAGTTGGCGGTTGGGCTGCAGCGTATGT-AAAGA 3291
Db 483 TGGATCTACACTAGTGGCGAGGAAGTTGGCGGTTGGGCTGCAGCGTATGTAAAGA 542
QY 3292 AGGATCCAGAAACTCGAGAGACCGGTATTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG 3351
Db 543 AGGATCCAGAAACTCGAGAGACCGGTATTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG 602
QY 3352 GGATATGCTGATCGATGAGTTCGACAAATGTCTGATAATGCCGGAAGCATGCTTCATG 3411
Db 603 GGATATGCTGATCGATGAGTTCGACAAATGTCTGATAATGCCGGAAGCATGCTTCATG 662
QY 3412 AGGTAATGGAGCAACAAACCGGTATCTGTAGCCAAAGGGGTATCATTTGCCCTGCTGA 3468
Db 663 AGGTAATGGAGCAACAAACCGGTATCTGTAGCCAAAGGGGTATCATTTGCCCTGCTGA 719

RESULT 4
LOCUS AX281540 723 bp DNA linear PAT 02-NOV-2001
DEFINITION Sequence 2 from Patent WO0177354.
ACCESSION AX281540
VERSION AX281540.1 GI:16608795
KEYWORDS Physcomitrella patens
SOURCE Physcomitrella patens
ORGANISM Physcomitrella patens
REFERENCE 1
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
costa e Silva,O.D., Bohnert,H.J., van Thielén,N., Chen,R. and Sarria-Millan,R.
TITLE Cell cycle stress-related proteins and methods of use in plants
JOURNAL Patent: WO 0177354-A 2 18-OCT-2001; BASF Plant Science GmbH (DE)
FEATURES location/Qualifiers
1..723
/organism="Physcomitrella patens"
/mol_type="unassigned DNA"
/db_xref="taxon:3218"
ORIGIN
Query Match 15.5%; Score 673; DB 6; Length 723;
Best Local Similarity 97.1%; Pred. No. 2.5e-189;
Matches 696; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 2753 GAAGCTGGAGATCGTATTGAGTTAACAGGAGTTTTCAGGCCATGGCAGTTCGAGTTGG 2812
Db 3 GCACAGGAGATCGTATTGAGTTAACAGGAGTTTTCAGGCCATGGCAGTTCGAGTTGG 62
QY 2813 TCCGAATCAACGAACATTACGAGCATTTGTATAAGACCTACATCGATTGCGTGACGTCAA 2872
Db 63 TCCGAATCAACGAACATTACGAGCATTTGTATAAGACCTACATCGATTGCGTGACGTCAA 122
QY 2873 GAAGTCTGACAGGGTTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGATGA 2932
Db 123 GAAGTCTGACAGGGTTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGATGA 182
QY 2933 TATGTATCTGGGTATCATGAAGTGATACCTTCAAGACTGCTATGAAGCAAGATTCA 2992
Db 183 TATGTATCTGGGTATCATGAAGTGATACCTTCAAGACTGCTATGAAGCAAGATTCA 242
QY 2993 AAAAATTAAAGAGCTGCTCAAGCTCCCGGCACTTATGATAGACTTTTCAAGTTCGCTGGC 3052
Db 243 AAAAATTAAAGAGCTGCTCAAGCTCCCGGCACTTATGATAGACTTTTCAAGTTCGCTGGC 302
QY 3053 TCCAGCAATTTGGGAGCTTGAAGATATTAAGAGGCTCTCTTTGCCAGCTCTTTGGTGG 3112
Db 303 TCCAGCAATTTGGGAGCTTGAAGATATTAAGAGGCTCTCTTTGCCAGCTCTTTGGTGG 362
QY 3113 GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGTGACATCAATGTTTACTTGT 3172
Db 363 GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGTGACATCAATGTTTACTTGT 3172
QY 3113 GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGTGACATCAATGTTTACTTGT 3172
```

Db 363 GAAGCTAAGAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGT 422

Qy 3173 TGGGACCTCGTACCACTAAATCTCAGCTGCTTCAAGTATGTGCAAGATAGCTCTCTCG 3232

Db 423 TGGGACCTCGTACCACTAAATCTCAGCTGCTTCAAGTATGTGCAAGATAGCTCTCTCG 482

Qy 3233 TGGATCTACACTAGTGGCGGAGGAAGTTCCGCGTTGGCTGACAGCGTATGT-AAAGA 3291

Db 483 TGGATCTACACTAGTGGCGGAGGAAGTTCCGCGTTGGCTGACAGCGTATGTAAACGA 542

Qy 3292 AGGATCCAGAACTCGAGAGCGGTATTTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG 3351

Db 543 AGGATCCAGAACTCGAGAGCGGTATTTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG 602

Qy 3352 GGATATGCTGTATCATGATGTTTCGACAAAATGCTGTATAATGCCCGAAGCATGTTTCATG 3411

Db 603 GGATATGCTGTATCATGATGTTTCGACAAAATGCTGTATAATGCCCGAAGCATGTTTCATG 662

Qy 3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATGCTCGCTGCGTGA 3468

Db 663 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGTTCATGCTCGTGAAGCCGA 719

RESULT 5

SPDC21

LOCUS

DEFINITION Schizosaccharomyces pombe cdc21 gene for protein required for S phase execution.

ACCESSION X58824

VERSION cdc21+ gene; cell division control; mcm4.

KEYWORDS Schizosaccharomyces pombe (fission yeast)

SOURCE Schizosaccharomyces pombe

ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.

REFERENCE 1

AUTHORS Coxon, A., Maundrell, K. and Kearsley, S.E.

TITLE Fission yeast cdc21+ belongs to a family of proteins involved in an early step of chromosome replication

JOURNAL Nucleic Acids Res. 20 (21), 5571-5577 (1992).

PUBMED 1454522

REFERENCE 2

AUTHORS Kearsley, S.E.

TITLE Direct Submission

JOURNAL Submitted (05-APR-1999) S.E. Kearsley, Oxford University, South Parks Road, Oxford, UK

REMARK revised by [3]

3 (bases 1 to 3667)

REFERENCE

AUTHORS Kearsley, S.E.

TITLE Direct Submission

JOURNAL Submitted (15-JAN-1999) S.E. Kearsley, Oxford University, South Parks Road, Oxford, UK

COMMENT On Jan 21, 1999 this sequence version replaced gi:4926.

FEATURES

source

1..3667

/organism="Schizosaccharomyces pombe"

/mol_type="genomic DNA"

/db_xref="taxon:4896"

641..3436

/gene="cdc21"

641..3436

/gene="cdc21"

/codon_start=1

/product="cdc21 protein"

/protein_id="CAA41628.1"

/db_xref="GI:4165293"

/db_xref="GOA:P29458"

/db_xref="UniProt/Swiss-Prot:P29458"

/translation="MSSSQSGRANELRTPGRANSSREAVDSSPLFPFASPSPGSTRLLTPRTARTPLASSPLFPSSPGPNIPQSSRSLLLSQRNDLFLDSSQRTPRSTRGDIHSSVQMTPRRREVDPQFQVSTPSSLLFSGSDALTFQAHPSSEVADTVRVIMGTVNSIQESIASFRGLRGFKKKRYRNLNMLPPDPAEQLVYIEALRNRMIMGLEIL

ORIGIN

Query Match 12.0%; Score 520.8; DB 15; Length 3667;
Best Local Similarity 59.5%; Pred. No. 1.2e-143;
Matches 980; Conservative 0; Mismatches 647; Indels 21; Gaps 5;

Qy 2380 GACCTTTCAATCTCAAAGCATCGGTGACATCGCGTGAACCTCAACCTTCAGATATAGACA 2439

Db 1482 GTCCTTTTAACTTGAAAAATGCAATTAATATGCGGTGATCTTAATCTCGTGACATTTGACA 1541

Qy 2440 AATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGG 2499

Db 1542 AATTGATCTCTATTAAAGGTCTGGTACTTAGGTGTACACCTGTAAATATCCCGCATGAAGC 1601

Qy 2500 GGGCCTCTTCAAAATGTTTAGTGTGTGCTACTCCCTCCGCTAGTTACAGTTGTTAAAG 2559

Db 1602 AAGCATTTTTTCGATGTTTCAGTTTGTGTGTCATTGTGTAC---CGTGAATTCGATAGAG 1658

Qy 2560 GCGCGTTTGAGGAGCCAAAGAGTGTGAAAAGCCAGAAATGTGCAGCAGCGAATGCTATGT 2619

Db 1659 GCGGAATGCTGAGCCCAATAAATGCCCTAGAGAAGTCTGTGTGCTACAAATGCAATGC 1718

Qy 2620 CTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGTCTTTCGAAGAACTC 2679

Db 1719 AGCTAATTCATAATCGTTCTGAATTCGCGCATAGCAAGTTATTAAACTCCCAAGACAC 1778

Qy 2680 CAGATGCCATCTCTGAAGAGAGACTCCACACAGATGAGTGTGTTATACACACTA 2739

Db 1779 CTGACGTGGTACCGGATGCCCAACCCACACACTCTGTAAGTCTGTGCGTATATGATGAAC 1838

Qy 2740 TGGTTGATGCTGTGAAGCTCTGAGATCGTATTGAGTAAACAGAGTGTTCGAAAGCCATCG 2799

Db 1839 TGGTTGATGCTGTGAGAGAGAGATCGTATAGAAAGTCACTGGGAATTTTCGCTGTGTAC 1898

Qy 2800 CAGTTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTTGTATAGACTCATCGATT 2859

Db 1899 CAGTTTCGTTGNAACCTCGCATGCGTCTGTAAATCATATTATAAACTTATGTCGACG 1958

Qy 2860 CGGTGCACGTCAAGAAAGTCTGACGGGTCGACTGCAAACTGAAAGTCTTATGGAGATGG 2919

Db 1959 TTGTTTACATATAAAAGCAGGACAAACGTCGTTGGGTAC---GGATCTCTTACGTTGG 2015

Qy 2920 ATAGAGGAATGATATGATGCTGGGTATCATGAAAGTCAATCTCAGAGCTGCTAAATG 2979

Db 2016 AAAGCGATATTGCTGAAGATGCTGCTTTAC---AAATTTGATGAAGTACGGAATAATCAG 2072

Qy 2980 AAGCAAGATTCAAAAACCTTAAAGAGCTGTCCCAAGCTCCGGGCATTTTATGATAGACTTT 3039

Db 2073 ACGAAGAAAGTGGAAAAAATTCACAAAGTTTCAAGCGGACGATATATACGATATTCTTT 2132

Qy 3040 CAAGGTGCTGCTCCAGCATTTGGGAGCTTTGAAGATATTAAAAAGGGTCTTCTTTGGC 3099

Db 2133 CCGGCTCTCTGGCACCCTAGCATTTTACGAATGATGATGTTTAAAGAGGGCTTTTCTTAC 2192

Qy 3100 AGCTCTTTGGTGGGAAGCTAAGAAAATTCATCTGGAGCATCT-----TTCGAGGTG 3153

Db 2193 AGCTATTGGTGGAGCAACAAGTCTTTTTCATAAAGCGCAAGTCCGAGATATCGTGGAG 2252

Qy 3154 ACATCAATGTTTTTACTTTGTTGGGACCCCTGGTACCGATTAATCTCAGTCTTCAGTATG 3213

Db	2253	ACATAAAATATCTTGATGTGTGGTGATCCATCTACGCTCTAAGTCTCAAAATTTTGAAGTATG	2312
Qy	3214	TGCAAGATAGCTCCTCGTGAATCTACACTAGTGGCGGAGGAGTTTCGGCGTTGGGC	3273
Db	2313	TCCATAAGATAGCACCCAGAGGTGTTACACTAGTGGCAAGGTTTCATCTCTGTGTGTT	2372
Qy	3274	TGACAGCGTATGTACGAAGATCCAGAAATCCGAGAGACGGTATTTGGAGAGCGGAGCTT	3333
Db	2373	TAACTGCCCTATATTAATCTCGTGACCAAGACACGAAGCAACTAGTCTTGGAAAGTGGTGT	2432
Qy	3334	TGGTCTTCTAGTATCGTGGGATATGCTGTATCGATGAGTTTCGACAAAATGCTGTGATAATG	3393
Db	2433	TGGTACTGTCAAGTGGGGCATTTCTGCTGCAATGATGAGTTTGACAAAATGCTCGATGCTA	2492
Qy	3394	CCGGAAGCATGCTTCATGAGTAAATGGAGGACCAACACGGTATCTGTACCCAAAGGGGTA	3453
Db	2493	CACGTTCTATTCTTCATGAAGTTATGGAGCAGCAACCGTTACAGTTTGCCAAAGCGCTGCA	2552
Qy	3454	TCATTGCCCTCGTGAACGCTCGGAGCTGTCTCTTGCAATGTCGCAATCCTAGTGGTCCC	3513
Db	2553	TTATTACTACTTTGAACGCTAGGACTTCTATTAGCTAGTGGCAATCAATTTGAAGCA	2612
Qy	3514	GATCAATGCGCGCTTTCTGTGATGATGAACATCCAGCTTCTCTCAACTCTACTTTCTA	3573
Db	2613	AATACAATCCAGACTTGCCTGTGCACCAAAATATTGATCTTCGCGCTACTCTCTTATCTA	2672
Qy	3574	GATTGATTTAATTTACTTAATGCTCGCAACACGAGACGACAAACGATCGTGTCTCG	3633
Db	2673	GATTGATTTGGTGTATTAATATTAGACCGCGTGGATGAACCGCTGGATAGGAAGCTCG	2732
Qy	3634	CCAGCATCTCTGTGCTTTTACACTATGAAA-----CTATGAAGTTTCAAGCAGGACG	3687
Db	2733	CAATCATATTTGTTCTATGATACATGGAAGATACCTCTGACATGCTACAGATATGGAGG	2792
Qy	3688	CCTTAGATCTACAAACACTTACCGGTATATCACCTATGCTCGTGCAGCATGTACATCTA	3747
Db	2793	TCTTTAGTGTGAATTTTGTGACAACTATATCACATATGCTAGAAATAATATTATCTCTG	2852
Qy	3748	CATTAAGTATGAAGCTGCTGAAGATTTGATTAATAGGCTATGTTGAGATGGCCCAAGG	3807
Db	2853	TCATTAGTGAAGAGGACGACAAAGGAACTAGTGAATGCTTATGTTGGTATGCGAAAGCTTG	2912
Qy	3808	GCAACTTCTCTGGAAGCAGTAAAGAGTGATAACAGCCACACCTCGGCAACTCGAAAGTA	3867
Db	2913	GGGAGATGTGGAGCATCTGAGAAAGATTAATCTGCTACACAAAGCAACTGGATCAA	2972
Qy	3868	TGATTGCTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTG	3927
Db	2973	TGATTGCTATCTCGAGGCCCATGCGAAATGTCATCTTCGCAACGTTGTTGAGGTAGGAG	3032
Qy	3928	ATGACAGCAAGACTGTGGCCCTTTTAGAGCTGCTTTGAGCAATCTCTGCTACTGATCATG	3987
Db	3033	ATGTTCTTTGAAGCTGCTCGCTTGATTAAGACCGCTATTAAGGATTTATGCAACTGATCTCG	3092
Qy	3988	CAACAGGTACGATAGACATGATCTTAT	4015
Db	3093	CAACAGGAAGATTTCTCTTGACCTAAT	3120

RESULT 6
AC116551/c

LOCUS AC116551 Dictyostelium discoideum chromosome 2 map complement (581327-427576)
DEFINITION strain AX4, complete sequence.
ACCESSION AC116551 AC116101 AC116549 AC116965
VERSION AC116551.2 GI:28829639
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 153751)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,J.P., Guigo,R., Kumpf,K.,

Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.
Sequence and analysis of chromosome 2 of Dictyostelium discoideum Nature 418 (6893), 79-85 (2002)
12097910
The Dictyostelium Genome Sequencing Consortium
2 (bases 1 to 153751)
Baumgart,C.
Direct Submission
Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
3 (bases 1 to 153751)
Baumgart,C.
Direct Submission
Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
4 (bases 1 to 153751)
Baumgart,C.
Direct Submission
Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
5 (bases 1 to 153751)
Baumgart,C.
Direct Submission
Submitted (21-PEB-2004) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
On or before Feb 21, 2004 this sequence version replaced gi:28850309, gi:19807752, gi:19920064, gi:19807784.
CDS predictions from GeneID do not necessarily reflect true genes.
Further information is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml)
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
Location/Qualifiers
1..153751
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="complement (581327-427576)"
/note="GeneID exon scores (in order of location ranges): 70.76, 0.70 - GSCJ_ID dd_02254"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52218.1"
/db_xref="GI:28829702"
/translation="MNESISDHNEKLFWKVFRNKYLNIIEPKQLESTIEYDSFTIAS LKLSVANNNDNNNNKLIIEFKLMKEKYLEGLDLVTISISSRNFELKILLNK PYSPLVETPDNALQKYGEISEKISLIQLIGNQININTFLIENNFOIKNGSPFKNT IQFTLNQNDNDNNKNNKLIKLELKIYSIKKDEILNLEIMONKLLDLAYM SKENFQNSIYWFKELYEFGVKTFIHLILLVFKIYKRSKSLNPNEIIEINIMF NSLLIKNEEKELKLLINETSQNLFLKLYIYRIEIECKLSKIVIRNLKTFNIF LIETSNNSNNKKEEIIIRIKNFVYEFKNNDKQINKFRVILNNLSKYYHK FTDTIKTYVYKRNKLIQFOSLELIEVAIKKLTNNNNNNNDNLPIKLETINWSM KLETIGNSQIKTFGLPIEWINLDDFLDDEINNKIIIKDQVTDWLSNMRNLNRI VYNNPKSIIDEFKQNESIINEDELNDKVTNFKTLKFEKNEFTILVOLIKYFIEK VLNKYGQDSYEILKNFKVDNLCCKFQKTELLELVSTINDLDMVILLGCGIDFTN QKDEILFNWHLITNTSTPIFYPIQINGININFKIITHIKFQYIINNKKWCEQTSLL ECKHLINHLFKKLIQTKTGLNLIYKIDFTINNNNNNNNNNNNNNNNNIILKLDHS KFHSIHYLNLSKPKLILYIESNQIRDWNVLLIPQ"
3515..5095
/note="GeneID exon scores (in order of location ranges): 144.31 - GSCJ_ID dd_02253"
/codon_start=1
/product="similar to Arabidopsis thaliana (Mouse-ear cress). Hypothetical 59.5 kDa protein"
/protein_id="AAO52156.1"
/db_xref="GI:28829640"

/translation="MNOVQPTLSNKAIEINRWLIDNGVRINNKLIKIVYLGKNNFE
QNTTATTTTSERLINDHISGLGVIISLKLKVDIVAKIPKSIILSHTSISNILE
KYIENNTGSIERLINDHISGLGVIISLKLKVDIVAKIPKSIILSHTSISNILE
VLNDDIILNOVADVIESILSKNHPFIPGKELYSIENKELANSISRAFCVDSYH
GDSVLPLADIENHQTARHVIHESNGDVNCKGSIKTKCHRVYVQHTVNSTKGRVT
HKVAGIPSSKHGNCSTYITINDEEDKOTIIEDEDEHYIKYKVGAVANKERT
TYGDHDAIENHAIENDRLDADALCVGIALAQSIQFQSWKSMSEHREKRYFQLEA
EDIVKQSVLVRNLTLSILNKLFSNYNNSNETTATTTTIBQDQELSTLKNQREIIAQSL
KICKETLKNKSNYKKNL"
complement (5235..7895)
/note="GeneID exon scores (in order of location ranges):
270.65 - GSCJ_ID dd_03185"
/codon_start=1
/product="similar to essential for initiation of DNA
replication; homolog of S. pombe CDC21; Cdc54p
[Saccharomyces cerevisiae]"
/protein_id="AAO52157.1"
/db_xref="GI:28829641"
/translation="MDPKETTKTSIFLNTTKNNNNNNNNNNNNNNNNNNNNNNNN
WNN
NGTTVKVDVXERPKFELNPTPKRERENNDINNNNNNNNNNNNNNNNNNNNNNN
DEBQDKLIKKEKVOYLRLDLKLVNFKCHLNINMKYLVQPMELVLOVSPENM
IPLDEELNLTRELPFPITPRNGDEDDDDDDDEDEDTRIELHFPNLLKTPMRDLN
PSDIDKIIISIGLIRSSSIPEIKOAFMCAVCEATFANVKGKIKQEPSECSNKS
KQSLSIHNRCLFGDKQYIKLOETPDALPEGETPHTVALFAYGDLIDIAKPGDRVELT
GVFKASPMRAGSNRLRSIYKTYIDILHKRTDKGYDDDDDDDDDDGGGTGTGKE
TGENLDJEDLDFDLSEKEREIIELSKKPDYIDIVTKSIAPNIWELEDIKKILCOL
FGGSKSVQDYGGKFRGINILLCDPGTSSKQLSYVHKIAPRGIYVSGKSSAVGL
TAYITKDPDTLESALVSDKGVCCIDFQNDQTRSLRHEVMEQQTYSIAKA
GLICTNARTSLASANPSGSRYPKLSVVENIQPPLLRSFDLIYLVLDKANERSD
ROLARHLSVMTWDETVPVSHFTPKETLNYIQYARKHINPKLITDLSAKLVQGYLEMR
SSGSKTKTSATPROLESIRIAEAHARIFSEFVEPLDVEAIRLIKVALQQAIDP
ENGTTDMDLITGSRASSREATRLKSHKIKGKKHLLTDQLLKLITQNVQVQIQ
IIEIIEALRQLQDESIHQSSGIDITYGVV"
join (8672..9027,9184..9463)
/note="GeneID exon scores (in order of location ranges):
37.11, 23.19 - GSCJ_ID dd_02251"
/codon_start=1
/product="similar to cytochrome c1 heme lyase
[Caenorhabditis elegans]"
/protein_id="AAO52158.1"
/db_xref="GI:28829642"
/translation="WGDNDNTNITSITKSPDTAMGGCPIVHDINPANHMYKPNQPH
PDQTPLESTRTITPTEKNDMOPSPQMFNMAKKQYEPKEDSMVTSINTV
NKKMEDVLQWENDKVCNPKLKFPGKATDFPKAKPLNTFLGYKLPPRHDWIV
DNKGKEVYDIYFGRINQSGKSIYIYDVRPAIDDFSSLKRLVHFFK"
complement (join (9650..13297,13390..13536))
/note="GeneID exon scores (in order of location ranges):
468.70, 11.48 - GSCJ_ID dd_02250"
/codon_start=1
/product="similar to C25a11.4b.p [Caenorhabditis elegans]"
/protein_id="AAO52159.1"
/db_xref="GI:28829643"
/translation="WNPPEMIQNIIVIKNAIEVGCIIIVNIGAQDILNATPYLVILGVI
WQIIRKGLSKVONANETLEILFEDSVKEVDNTQNSDSEAEILLRWVNYHLEK
EGCERINSESDIQDSVYVSHLPQVPIEFENLVSTIHSNLSFRAELITKACEK
IVGKPLPSDIALGPKNLALVASFNSEAIAINREREREKERTDOERLERERLE
RERIRERLERERLEQERLERERLEQERLERERLEQERLERERLEQERLERER
LEQERIRERLER
RER
EYKICIRIQRMIDTSLDNTMKLIGDITELKNITIGELRNTHHQELKYLKTSID
RVLENKQKKKKQ
LDNQKQ
MKLLYQYFKKKQKYLKEIITPLIKVLQDITLNPFGYIWKETIQIAELKDGKGS
EKKTLDQTEQLMANNVVSIFNSRCLQVLEGNAPLERITKSIEOLPVQLRYLCKRII
AKSKYAFPTNNNNHNGGGISSDESGSTANIITNNRSKOLPENREKYISYFLII
YRPIIIPSSDLNELLVLQHPYITKQSRFNLSQLSRLIQSVFSEKPMWYPTVYL
NWKIIONSSADKFLCQVQVFDLDELHYSEYNEIMLENNIKFCSVINSEIILWHE
TILSNFYDLIEDLTKEVSEKSIQNLAEARINSKLDKGYVTAARQSPAIMTSQSKQ
NNPLKKYDEIDPFIQPEQGDRIQSLSPFNRLDQLEQSIDLWKECKLMDILSKS
QPPKQVNTIIEIRIADYFKTGSNSSENLSIINKSNELISRLRTFDRDQCTDGLDGG

CPDNFNLFLDVEKELATILRKNKMRSELTRLHSLKSIIRIDAAPLEHLTKFQNEQL
EKALVQKLTQTLPFKSTIKPHKFIILLDLVKGMILRCLPSSPSFKTKIYIK
SLGGLPEIRAEKFGPITKSLQLEDDLLERKSNHIDBPPLEGITLDDVTTLTHLMLKF
SL"
complement (join (13685..13812,14033..14207))
/note="GeneID exon scores (in order of location ranges):
6.21, 14.57 - GSCJ_ID dd_02249"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52160.1"
/db_xref="GI:28829644"
/translation="MTYDTSRYNTLRGSKMDPPFITQEVLASAKISKKNILINAN
ABDSVHTFSEEEKVAVSKPITERLIDQEDVLSVNPIDPSTQDLFVNTDGVLLK"
16650..17525
/note="GeneID exon scores (in order of location ranges):
78.29 - GSCJ_ID dd_00829"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52161.1"
/db_xref="GI:28829645"
/translation="METKDKKKMELLGDRMKSYEDDMKIQIERNKPFIRLDGHSFS
KFTNFKNPHDIRHNAMIESTVLLKTFMPTCIYTFSEITMCPFSIDESLEBGEKE
IPNLAYSQVOKLISLSSGLASTVFFKSIITNAQYDKDTLNLKLETCPTPHDARIF
TLPNQEIYVNNLIWRLSLVCKRNSVANLGAHFTPKOMGLNNTVRKKLLEKSIY
ENEPAWYKPYLKKKYYTTLTVSPNEPEKQITAIRSKVNCESFPIITKLSNLSNISP
KTLPEGFNLELIESK"
complement (join (17735..18959,19168..19640,19750..20190,
20286..20531))
/note="GeneID exon scores (in order of location ranges):
130.29, 48.89, 51.03, 20.75 - GSCJ_ID dd_00831"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). ALG-2 interacting protein X"
/protein_id="AAO52162.1"
/db_xref="GI:28829646"
/translation="WLSIERKTEKVFSPKPLTKYIKQFSKAESDQHETOIATLNLG
REDVNLQERTSEKEMWVKYISLSSLELFPPISENVRISFPPTDSVRQKSTLYS
IYFERASVLPNGYSIVQIASSGNRSIEGKACNQLQAAAGVANKUREYASLHPEC
STSDAFSSQALVITMLAQECIYKASMDNLSLSKLAQVAYDYDTFNQLL
NSNLSKIVDRNNWITATVSKYLAISYCHAGLEQVQSGEOVQSLMIAVDNINO
SKVNLAKTAPIELKEIVERYVTSIYCESAKKONDITVHDTIPAHKLPTPEKKPLA
KALPLPEINFDVPPNSLVPFSPVSKEDSAYNDQKELLRLKELDNIFHNQSAKASLLSM

Query Match 11.3%; Score 492.4; DB 2; Length 153751;
Best Local Similarity 59.3%; Pred. No. 6.6e-135;
Matches 1011; Conservative 0; Mismatches 636; Indels 57; Gaps 8;
QY 2370 ATTGAGGCCAGACCTTCAATCTCAAGCATCGGTGCACATCGGTGAACCTCAACCTTCA 2429
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 7121 ATTGAATTACATCCATTCAATTTATTAGAAAAAACCAATGAGAGATTTAAATCCATCA 7062
QY 2430 GATATAGACAATGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCT 2489
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 7061 GATATTGATAAGATTATATCAATTAGTGTGTTAAATTTATTAGAGTAGTTCATTAATACCA 7002
QY 2490 GAAATTAAGGGGGCGCTTCTTCAATAGTTTGTGTGTGTCTACTCGCTCCGCTAGTTACA 2549
Db 7001 GAGATTAACAAGCGTTTATGATGTGCAGTTTGTG---AAGCAACTTTTTCATGCAAAAT 6945
QY 2550 GTTGTAAAGGGCGGGTGTAGAGCCCAACAAGGTGTGAAAAGCCAGAAATGTGCACACGG 2609
Db 6944 GTTGAGAAGGGTAAAAATTCAAGAGCCATCA-----GAATGTTTCAAAATTTAAATCAAAAG 6891
QY 2610 AATGCTATGCTCTTATTACATCATGTCACCTTTTGGCAATAAGCAGATAGTGCCTCTT 2669
Db 6890 CAATCATTTCAATCATTCATTAATCGTGTGTTGTTGGTGTAAACAGTATATTAACCTT 6831
QY 2670 CAAGAACTCCAGATGCCATTCTTGAAGGAGAGACTCCACACACAGCTCAGCATGTGTTTA 2729
Db 6830 CAGAAACACAGATGCAATTCAGAGGGTGAACACACACATACCGTCGCATTTATTCGC 6771
QY 2730 TACAACACTATGTTGATGCTGTGAAGCTCGAGATCGTATTGAGGTAAACAGAGTTTTC 2789
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 6770 TATGCTGATTAAATGATATTGCAAAACACAGGTGATCGTGTGAGTAACTGGTGTATTT 6711
Qy 2790 AAGGCCATGGCAGTTTCGAGTTGGTCCGATCAACGACATTTACGAGCATTTGTATAGACC 2849
Db 6710 AAGCAAGTCCAAATGAGAGCTGGTTCAAAT---AGATCACTTTAGAAAGTATTTATAAAACT 6654
Qy 2850 TACATCGAATGGCTGCACGTCAAGAAGTCTGCAGGGGTGCACTGCCAAACTGNAAGATCCT 2909
Db 6653 TATATTGATATTTACATATATAAAGAACAGATAAAGGTAAATATATGATGATGATGATGAT 6594
Qy 2910 ATGGAGATGGATAA-----GGAGAAATGATATGATGCTGGGTATTCATCAAAAGTGATACT 2963
Db 6593 GATCATGATGATAAATFACTGGTGGTGGCGACAGGTACAGGTAAAGAAACAAATGARAAT 6534
Qy 2964 TCAGAAAGCTGCTAA-----TGAAAGCAAGATTCAAAACTTTAAA 3002
Db 6533 TTAGATTTTGAAGATTTAGATGAATTTGATTTTATCTGAAGAAAAAGAAAGAGATTTATA 6474
Qy 3003 GAGCTGTCCAAAGCTCCCGGCAATTTATGATAGACTTTTCAAGTGTCTGGCTCCAGCATTT 3062
Db 6473 GAATTTATCAAAAAAACAGACATTTATGATATTGTAATTAATCAATAGCACCAAAATATT 6414
Qy 3063 TGGGAGCTTTGAAGATTAATAAAAGGCTCTTTTCCAGCTCTTTGGTGGGAAGGCTAAG 3122
Db 6413 TGGGAATTAGAAGATTAATAAAAGGTATACCTTGTCAATTTATTTGGTGGTTCAAAAAG 6354
Qy 3123 AAAATTCCA---TCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTTGTTGGGAC 3179
Db 6353 AGTTATCAAGATTTATGGTGGAAAGTTTAGAGGTGATATAAAACATTTTATTTGTGTGAT 6294
Qy 3180 CCTGTGTACAGTAAATCTCAGCTGCTTCAGTATGTCACAGATAGCTCTCTCGTGGATC 3239
Db 6293 CCTGTGTCTCGAAATCTCAATTAATCTCGTATGTTTCAAAATTTGCAACCCGCTGGTATC 6234
Qy 3240 TACACTAGTGGCGCAGGAAGTTTCGGCGTTCGAGGTGACAGGTATGTAACGAAGATCCA 3299
Db 6233 TATACAAGTGGTAAGGGTAGTAGTCAGTTGGTTGACGGCATATATCACCAGATGCC 6174
Qy 3300 GAACTCGAGAGACGGTATTTGAGAGCGGAGCTTTGGTTCCTTGTAGTATCGTGGGATATGC 3359
Db 6173 GATACAAAGAGACCGTTTTAGAGAGTGGTGCATTGGTACTCTCTGATAAGGGTGTGTTGT 6114
Qy 3360 TGTATCGATGAGTTCGACAAATGCTGATTAATGCGCCGAGCATCTTCATGAGTATATG 3419
Db 6113 TGTATTGATGAATTTGNTAAAAATGAATGATCAAACTCGTTCAATTCCTTCATGAAGTTATG 6054
Qy 3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTTGCTCGCTCAACGCTCGGACG 3479
Db 6053 GNAACAGCAACAGTATCAATTCGAAAGGGGGCATTTCTGCACGTTGAAATGCCAGAAC 5994
Qy 3480 TCTGTCTTGCATGTGCAAACTCTAGTGGGTCCGATCAATATGCGCGCTTTCTGTGATTT 3539
Db 5993 TCGATTCTTGCACTGCAAACTCAAAGTGGTTCACGCTTATATGCCCAAAATTCAGTGGTG 5934
Qy 3540 GATAACATCCAGCTTCCTCCAACTCTACTTCTAGATTTGATTTAAATTTACTTAATGCTC 3599
Db 5933 GAGAAATATCCAAATTCCTCCCAACCTTATATCGAGATTCGATTTAAATTTATTTAGTATG 5874
Qy 3600 GACAAACAGACGAGCAAAACGATCGTCTCGCCAGGCATCTCGTGGCTTTTACATCAT 3659
Db 5873 GATAAAGCAATGAAGAAGTGATCGTCAATTTGGTCTGTCATTTAGTATCGATGTATTTGG 5814
Qy 3660 GAAACTATGAAGTTTCAAGACGAGCCCTTATGATCTTACAAACACTTTACCGCGPATATC 3719
Db 5813 GATGAAACACCGGTTTC---TCAATTTCAATCCCAAGGAAACCTTCACAAACTATATT 5757
Qy 3720 ACCTATGCTCGTACATGTATCTTACATTAAGTATGATGAGCTGCTGAAGATTTGATTT 3779
Db 5756 CAATACGCTAGAAAAATATTAATCCAAAGTTAACTGATGATCTTCGCAAAATGTTTAGTT 5697
Qy 3780 AATGCTATGTTGATGCGCCAAAGGCAACTTTTCTCGGAAGCAGGTAAAGAGGTGATA 3839
Db 5696 CAAAGTTATTAGAGATGAGAGTATGGT-----AGTAGTAAAAAACAATC 5649

Qy 3840 ACAGCCACCTCGGCAACTCGAAAGTATGATTCGTATCACTGAGCCCTAGCTCGAATG 3899
Db 5648 TCTGCCACACCAAGCAATTTGGAATCACTTATTCGTATCGCTGAAGCTCATGCTAGAAAT 5589
Qy 3900 AGATTTTCTGAAGTGGTAGAGAAAGTTGATCGACAGAAAGCTGTGCGCTTTTACAGCTC 3959
Db 5588 CGTTTCTGNAATTTGTTGAACCATTAGATGTTGGAAGAGCAATTCGTTTAAATTAAGTC 5529
Qy 3960 GCTTTGCAAGCAATCTGCTACTGATCATCAACAGGTACGATAGACATGATCTTTATCAG 4019
Db 5528 GCACCTTCAACAAGCTGCAATCGATCCTGMAAATGGTACAAATCGATATGATCTAATTACA 5469
Qy 4020 ACTGAGTGTCCGCCAGGAGCGT 4043
Db 5468 ACTGCTGTTCTGCAAGTTCTCGT 5445

RESULT 7
AX489180
LOCUS AX489180 2733 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 6480 from Patent WO02053728.
ACCESSION AX489180
VERSION AX489180.1 GI:22323192
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
1
REFERENCE Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlssen, K. L.
AUTHORS Gene disruption methodologies for drug target discovery
TITLE Patent: WO 02053728-A 6480 11-JUL-2002;
JOURNAL Elitra Pharmaceuticals, Inc. (US)
FEATURES
source 1..2733
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

ORIGIN
Query Match 11.0%; Score 480.2; DB 6; Length 2733;
Best Local Similarity 56.6%; Pred. No. 1.7e-131;
Matches 959; Conservative 0; Mismatches 718; Indels 18; Gaps 3;
Qy 2409 ATGCGTGAACCTCAACCCCTTCAGATATAGACAAATTTGGTTCTCTGTTAAAGGAATGGTTATC 2468
Db 883 ATTCTGNAATTTGATCTTAATGATATTGATTAAGTTAGTTAGTTAAAGGGTTAACAATTA 942
Qy 2469 CGGTGCAAGTCTCATACACCTGAAATTAAGGGGCGCTTTCTTCAAAATGTTTATGTTGTTGGT 2528
Db 943 CGTTCTACTTCTATCATTTCTGATATGAAAGTGGCAATTTTTCAGATGTAACGCTGTGCT 1002
Qy 2529 CACTCGCTCCGCTAGTATTCAGTTGTTAAAGGGCGGTTGAGGAGCCCAAGGTGTA 2588
Db 1003 CATACTGTGGG---GTGGAAATTTGATCGTGGTGTAAATTTCCGAACCAACAAATGTCT 1059
Qy 2589 AAGCCAGAATGTGCAGACACGGAATGCTATGCTCTTATTTCAATTCGATCGACTTTTGTCA 2648
Db 1060 AGAGAAGTATGTGGTCAACCAATTCATGTGACTTATTCATATATCGTTTCATCTTTCT 1119
Qy 2649 AATAAGCAGATAGTGGCTTTTCAAGAACTCCAGATGCCAATCTCTGAAGGAGAGACTCA 2708
Db 1120 GATAAACAAGTGAATTAATTAACAAGAACTCCTGATTTGGTACCTGATGTGTAACCCCT 1179
Qy 2709 CACACAGTCACTGTGTTTATACAACACTATGTTGATGCTGTGAGCCCTGGAGATCGT 2768
Db 1180 CATTCTATCAATTTATGTGTTTACGATGAATTTGGTTGATCTTGTGCTGGTGGTATCGT 1239
Qy 2769 ATTGAGGTAAACAGGAGTGTTCAGAGCCATGCGAGTTTCGAGTTGGTCCGAATCAACGAACA 2828
Db 1240 GTTGAAGTATGTGTTATTTTCAGATCAACTCCAGTTCGAGCAAAATCCTTAGACAAAGACT 1299

Qy	2829	TTACGACATTGTATTAAGACTTACATCGATTGCGTGCAAGTCTGAAGAGTCTGACGGGT	2888
Db	1300	TTAAAAAATTATATAAAACATATTTGGATATTTGTTCTATGCTGAAAAAATTCGATAAAGG	1359
Qy	2889	CGACTGCAAACTCGAAGATCCTATGGAGATGATAAGGAGAAATCATATGTATGCTGGGTAT	2948
Db	1360	CGATTAGTGGTGATGTTACCA-----CTTTAGAACACGAAATTTGGCTGAAAAGGAT	1410
Qy	2949	CATGAAAGTGATACTTCCAGAAAGCTGCTTAATGAAGCAAAAGATTCAAAAAATTAAGAGCTG	3008
Db	1411	CAAGAAGTTGAAACAAAGTTTCGAAAAATCCTCTGCTGAAGAAGAAGCTAAAAATCAAGAATA	1470
Qy	3009	TCCAGCTCCGGGCATTTATATGATAGACTTTTCAAGGTGCGTCTGAGCTCAAGCATTTGGGAG	3068
Db	1471	TCTGAACGTGATGATTTGTATGAATAATTTGGCTAGATCATTTAGCTCCATCTATTTATGAG	1530
Qy	3069	CTTGAAGATATTAAGAAGGTCCTTTTGGCCAGCTCTTTGCTGCGAAGGCTCAAGAAAAAT	3128
Db	1531	ATGGATGATGTTAAAAAGGGGATCTTGTTACAATTTATTTGGTGATCTAATAAACAATTT	1590
Qy	3129	CCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTTTTACTCTGTTGGGAGCCCTGGTACC	3188
Db	1591	ACAAAAGTGGTCGTTATAGAGGTGATATAATTTATTTATTTGTTGGTGATCCATCTACT	1650
Qy	3189	AGTAAATCTCAGCTGCTTCAGTATGTCACAAGATAGCTCCTCGTGGAAATCTACACTAGT	3248
Db	1651	TCTAAATCACAATCTTTACAATATGTTCAATAAGATTGCTCCAAGAGGGGTTTATCTTCT	1710
Qy	3249	GGGCGAGGAAGTTCCGGCGTTGGGCTGACAGCGTATGTACGAAGNATCCAGAAATCTCGA	3308
Db	1711	GGTAAAGGTTCTTTCGGCTGTTGGTTTGACAGCTTATATCACAGAGANATTTGATACAAA	1770
Qy	3309	GAGACGGTATTTGGAGAGCGGAGCTTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT	3368
Db	1771	CAATTAGTATTTGAAAAGTGGGGCCCTGTGTATCTGATGAGAGGTGTTGTGTATGAT	1830
Qy	3369	GAGTTCCGAAAAATGTCTGATPAATGCCCCGAAGCATGCTTCATGAGGTPAATGGAGCAACAA	3428
Db	1831	GAGTTTGACAAAGATGAGTGATGCCACAGATCAGTTTTTACATGAAGTCATGGAACAACAA	1890
Qy	3429	ACGGTATCTGTAGCCAAAGGGGTATCATTTGCCCTGCTGAGCGCTCGAAGTCTGTCCTT	3488
Db	1891	ACTATTTCTATTGCTTAAAGCGGTATTTATTACCACATTTAAATGCTAGGACATCTATTTTA	1950
Qy	3489	GCATGTCAAAATCCTAGTGGGTCCCGATCAAAATGCGCGCCCTTCTGTGATTTGATAACATC	3548
Db	1951	GCCTCTGCAAAATCCAAATTAATCTCGTTATGATCCAAATTTGCTGTGACGGGTACATTT	2010
Qy	3549	CAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTTAATGCTCGCAAAACCA	3608
Db	2011	GATTTGCCACCACCATTTGTTATCACGTTTTGATTTGGTTTTATTGATTTTGGATAAAGTT	2070
Qy	3609	GACGACGAAAACGATCGTCTCTGCCAGGGATCTCTGGCTTTTACACTATGAAAACATAT	3668
Db	2071	GATGAATCAATTCGATAGACAAATTTGGCTAGACATTTGACTGATATGTTATTTGGAGATGCT	2130
Qy	3669	GAAGTTTCARAGCAGGACGCTTAGATCTAC-----AAACACTTACCGGATATATCACC	3722
Db	2131	CCTGAAACTGTTAATGCCAAATTCGGTATTTACCGTGGAAATTAATAACACTTTATTAATCAA	2190
Qy	3723	TATGCTCGTCAGCATGTACATCTCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAAT	3782
Db	2191	TATGCTAAAGAGATTTCAATCCCGTGTATGACAGAGAGGTAAAAATGAATTAGTTAGA	2250
Qy	3783	GGCTATGTTGATGCGGCCAAAAGGGCAACTTTTCTGGAAGCATGTAATAAAGGTGATAACA	3842
Db	2251	TCATACGTTGAAATGAGAAAAATTTGGGGAGGATGCTAGATCATTCAGAAAAAGAAATCACC	2310
Qy	3843	GCCACACTCGGCCAATCTGAAAGTATGATTTGATATCAGTGAAGCCCTAGCTCCGAATCAGA	3902
Db	2311	GCTACTACAAGACAAATTTGAAATTCGATGATTAAGTTGTTCTGAAGCTCATGCTTAAATGAGA	2370
Qy	3903	TTTTTCTGAAGTGGTAGAAAAAGTTTATGTCAGCAGAGAAGCTGTGCGCCCTTTTAGACGTGCT	3962

Db	2371	TTATCTGAAGAGTGTGAATTCATCGATGTTAAAGAACGAGTAAGATTAATTAATCGGCC	2430
Qy	3963	TTGCAGCAATCTGCTACTGATCATGCAACAGAGTACGATAGACATGATCTTATCAGCACT	4022
Db	2431	ATTAAAGATTATGCTACTGATCCAGTTACTGGTAGNATCGATATGATATGATTTCAAACT	2490
Qy	4023	GGAGTGTGGCCAGCAGCGATTTCGTGGGCGCACTTGTCTAGCTGCTCTGCGAGAGCTT	4082
Db	2491	GGTACTACGGCACAAAGAAAGAGTACAAGAAAGATTGGTCAGTGAATTTATGAAATC	2550
Qy	4083	ATACGAGTAAATTT 4097	
Db	2551	ATCGAAGAAATAAT 2565	

RESULT 8
 CR954214 1/c
 WPCOMMENT
 Sequence split into 6 fragments LOCUS CR954214 Accession CR954214
 Fragment Name Begin End
 CR954214_0 1 110000
 CR954214_1 100001 210000
 CR954214_2 200001 310000
 CR954214_3 300001 410000
 CR954214_4 400001 510000
 CR954214_5 500001 516206

Continuation (2 of 6) of CR954214 from base 100001 (CR954214 Ostreococcus tauri strain
 Query Match 10.8%; Score 471.4; DB 14; Length 110000;
 Best Local Similarity 57.0%; Pred. No. 1.2e-128;
 Matches 968; Conservative 0; Mismatches 681; Indels 48; Gaps 4;

Qy	2380	GACCTTTCAATCTCAAGATCGGTGCAATCGGTGAATCAACCTTCAGATATGACA	2439
Db	72194	GCCCGTTCAACATGATGAGCGCAACCCATCGCGACTTGAATCCATCGACATCGACA	72135
Qy	2440	AATTGGTTCTGTTAAAGATGGTTATCCGGTGCAGTTCTATCATACCTGAATTAAGG	2499
Db	72134	AGATGGTGTGGTGGTGGTATGGTGACGCGATGCAACAGCATCATCTCGACCTAAAC	72075
Qy	2500	GGGCGCTTCTCAATGTTTAGTGTGTGGTCACTCGCGCTCGGTAGTTACAGTTGTTAAAG	2559
Db	72074	TCGCGTATTCAAGTGTCTGATGTGGCGTTTCGCCCGAGAACATATTCAGTGCACGCG	72015
Qy	2560	GGCGGTTGAGGAGCAACAGGTGTGAAGAGCCAGAAATGTGACGACCGAAATGCTATGT	2619
Db	72014	GTGCGTCAACGAACTTC--GCTCAAGTGCACAGAGTGGCGCAAGCCAGGAACGATGA	71958
Qy	2620	CTCTTATTACAAATCGATGCACTTTTGGCAATTAAGCAGATAGTGCCTTTCAAGAAATC	2679
Db	71957	CTTGATTCAAAATCAATGCGTCTTCGCAACCAAGCAACGGTGAAGATGACAGAGACAC	71898
Qy	2680	CAGATGCCATTCTGAAGGAGAGATCCACACACAGTACGATGTGTTTATACAACTA	2739
Db	71897	CGGATCGATTCCCGAGGAGTGAAGTCCCGACACTGTATCCATGTGCGTCTCGATGATC	71838
Qy	2740	TGGTTGATGCTGTGAGCTCGAGATGATTTAGGTTAAGGTTAAGAGTTTCAAGGCCATGG	2799
Db	71837	TCGTTGATCAGCGCAAGCCAGGTGACCGAGTTGAAGTTACTGAGATGTATCGTGGCGTTC	71778
Qy	2800	CAGTTTCGAGTTGGTCCGATCAACGAATTAAGCAGTATGATAGACCTACATCGATT	2859
Db	71777	CAATTCGCTGAGCTCTACGAAGCGCACGCTCAAGAGTGTGTTACAGAGCTACCTAGACG	71718
Qy	2860	GCGTGCAGCTCAAGAAGTCTGACAGGGGTGCACTGCAAACTGAAGATCCTATGAGATGG	2919
Db	71717	TTTTTCATATAAGAAGGAGCTGCGGGCGAGGATGCGAAACCGCGGACCCGAGGATG	71658
Qy	2920	ATAAGGAGAAATGATGATGTCG-----GGTATCATG	2952
Db	71657	AAGAGGCTGCAAGAACTCAGCTGCTTCAACGAAGTCACTGGAACGGGTGAAGAACTCAAG	71598


```
QY 2953 AAAGTGATATCTTCAGAGCTGCTAATGAAGCAAGATTCAAAACCTTAAAGAGCTGTCCA 3012
Db 71597 GTCTGGGACAGCAAAATGGAGTTCACTCCGGAGCGCATGCCGAGATAGAGAGCTCGTGC 71538
QY 3013 AGCTCCGGGCAATTTATGATAGACTTTCAAGTGCCTGCTCCAAAGCAATTTGGGAGCTTG 3072
Db 71537 GTTCCCGGACATTTATGACCGGCTTGTGGCGCTGTGGCCCGTCAATTTGGGAGTTAG 71478
QY 3073 AAGATATTAAAGAGGCTCTCTTCCAGCTCTTTTGGTGGGAGGCTTAAGAAAATTTCCAT 3132
Db 71477 AAGAGCTTTAAAGAGACTTGTGTGTCAGCTCTTTTGGCGCCACGAACAGAGCTTCACGG 71418
QY 3133 CTGGAGCATC-----TTTCCAGGTGACATCAATGTTTACTTCTGGGACCCCTGGTA 3186
Db 71417 ACAAGGACGAACAAGTCCGGGGGACATCAACATTTTACTCTGCGTGTATCCGGGG 71358
QY 3187 CAGTAATCTCAGCTGCTTCAGTATGTGACAAAGATAGCTCTCTCGTGGAACTTACACTA 3246
Db 71357 TCGCCAAGTCTCAGCTATTGACGTACGTTTCATCGTATCGCGCGCTGATGTACACCT 71298
QY 3247 GTGGGCGAGGAAGTTCCGGGTTGGCTGACAGCGTATGTAACGAAGATCCAGAACTC 3306
Db 71297 CAGGTAGAGGCTCGTCTCGGTTGTCTTTACAGCGTACGTCACTCGCGATCGGAGTCAA 71238
QY 3307 GAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTTAGTGATCGTGGGATATGCTGTATCG 3366
Db 71237 AAGACATGTTGCTCAGTCAGGACGGCTGCTGTTTCTGATCGCGCAATTTGTTGCATTG 71178
QY 3367 ATGAGTTCGACAAAATGTCTGATAATCCCGAAGCATGCTTCATGAGTAATGGAGCAAC 3426
Db 71177 ATGAGTTCGACAAAGATGTCTGACAGTGTGATCAATGTTGCACGAAGTATGGAGCAGC 71118
QY 3427 AAACGGTATCTAGCCAAAGGGGTATATGCTCGCTGACGCTCGGAGCTGTGTC 3486
Db 71117 AGACGGTTTCCATCGCAAGGCTGGAATCAATTGGCGGTGTAAATGACCGGACTTCGGTCC 71058
QY 3487 TTGCATGTGCAAAATCTAGTGGGTCCGCATACAATGCGCGCTTCTGTGATTGATAACA 3546
Db 71057 TTGCTGACGNAACCGAGTCGGAAGTGTGTACACCGCAACATGTCATGTTGGAGAAAC 70998
QY 3547 TCCAGCTCTCTCCAACTCTACTTTCTAGATTGATTAAATTTACTTAATGCTCGCAAAAC 3606
Db 70997 TTCAACTCCCGCGACACTGTTGTCCCGTTTGTGTTGTTGCTTACCTAATGCTTGACCGTC 70938
QY 3607 CAGACGACGCAAAACATGCTGCTCTCGCAGGCACTCTGTCGCTTTACACTATGAAACT 3666
Db 70937 CTAATCCCGAGACGGATCGCGCTCTGGCGAGACACTTAGTCAGTTTGCACACTACAGAAATC 70878
QY 3667 ATGAAGTTTCAAAGCAGGACGCGCTTAGATCTACAAACACTTACCAGCTATATCACTATG 3726
Db 70877 CCCCAGAGAAACGGGGCGTGATTTCTGCCGACTTCTCACCAGTACGTTCTCGTATG 70818
QY 3727 CTGTCAGCATGTATCATCTCATTAAGTGTAGAGCTGCTGAAGATTGATTGAATGGCT 3786
Db 70817 CTGCGGCAATGTACAAACCGGTGCTCAGCAGCAAGGCTCGGAGAGCTCGTGGAGGAT 70758
QY 3787 ATGTTGAGATGGCCAAAGGCAACTTCTCGGAAGCAGTAAAGGTGATAACAGCCA 3846
Db 70757 ATGTTGAGATGGCGCATG-----GGTGGTAGTTCGAAAAGTTATTACTGCAA 70710
QY 3847 CACCTCGGCAACTCGAAAGTAGTATTGCTGATCAGTGAAGCCCTAGCTCGAATGAGATTTT 3906
Db 70709 GCGGAGAACAACTTGAGTCTCTCATTCGTCTCTCAGAGTCTCTAGCGCGCATGCGGTTTAA 70650
QY 3907 CTGAAGTGGTAGAAGATTGATGACGACAGAGCTGTGCGCCTTTTAGACGCTCGCTTTCG 3966
Db 70649 GTGCGGTGTGTGATCGTGACGACGCTAAGGAGCGCTCGGATTGATCGGGTGGCGATGC 70590
QY 3967 AGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGATCTTATCAGCATGGAG 4026
Db 70589 AACATCCGCGCTTGTCTCTCGCACTGGGACGATTGATATGATAAGATATTGACCGGAC 70530
QY 4027 TGTCGGCCAGCGAGCGT 4043
```

Db 70529 ATAGCGCCAGCGACCGT 70513

|||||

RESULT 9

SCU14731

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Saccharomyces cerevisiae

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 5392)

Whitebread, L.A. and Dalton, S.

Cdc54 belongs to the Cdc46/Mcm3 family of proteins which are

essential for initiation of eukaryotic DNA replication

Gene 155 (1), 113-117 (1995)

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

SUBMITTED (15-SEP-1994)

Regulation, Roche Institute of Molecular Biology, 340 Kingsland

Street, Nutley, NJ 07110-1199, USA

Location/Qualifiers

1. 5392

/organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"

/db_xref="taxon:4932"

1921. 4722

/gene="CDC54"

1921. 4722

/gene="CDC54"

/codon_start=1

/product="Cdc54p"

/protein_id="AAA86310.1"

/db_xref="GI:608171"

/translation="MSQQSSSPTKEDNNSSPVVNPDPSPQLSSPALFYSSSSSQG

DIYGRNNSSQLQVHKITPRGVTYSGSGSAVGLTATITRDVDTKLVLSGSAVLVS

RNRVHSLRSDRALPTSSSIRNGQNRVHMRNDIHTSPRI VDFDTRSGVNT

LDTSSAPSEASEPLRIWGTNVSIQECTNFRNFMLSFKYKRLDREERFINN

TDBELIYIKQINREMLGNSLNDANLLAYKQTELYHQLNLPOEVISIMDQTI

KCMVSLIVDNLDYDDEITBTFYKVRYPYNGVSKGMRLEPNPDIDKLNILGLVLR

STVPIPMKVAFFKNCVDHTMAYEIDRGVIOEPACERIDCNENSMILHNRCSFA

DKVILQETPDFVPGDTPHSISLCVVDLVDSCRAGDRIEVTCTFRSIPITRANSRQ

RVLSKLYTVVHVVKVSKRLDVTSTIEOLMKNKVDHNEVEVRVQITDQDLAK

IREVAAREDLYSLLARSAPSIYELEDVKGILLQLFGGNTKFTFKGGRYGDINILL

CGDPTSKSLQILQVHKITPRGVTYSGSGSAVGLTATITRDVDTKLVLSGSAVLVS

DGVGCCIDEDFKMSDSTRSLVHEVMOQTISIAGAGIITTLNARSSILASANPIGSR

NPNLPVTENIDLPPLLSRFDLVLDKDEKNDRELAKHLTNLYLDEKPEHISQDD

VLPEFLTMTYSYAKSHIHPITTEAAKTVELVRYVGMKMGDDSDRSDEKRTATTTRQL

ESMLRLAEAAKMKLVNVELEDVQEAURLISAKVDATDPTKTKIDMNLVOTCKSV

IQKRLQDLSREIMNVLDQASDSMFPNELIKQINHSQDRVESDDIQEALSRLQED

KIVLIGSGVRRSRVRLNRRV"

ORIGIN

Query Match

Beat Local Similarity

Matches 948; Conservative

10.8%; Score 470.8; DB 15;

Pred. No. 1.2e-128;

Length 5392;

0; Mismatches 74;

Indels 9;

Gaps 2;

QY 2364 AAGCATATTAGGCCAGACCTTTCAATCTCAAGCATCGGTGCACATCGTGAATCAAC 2423

Db 2815 AAGTTTTATAAGTAAGCCCTTACAAATGTAGTCTTCTGTAAGAGGTATGCGCAATTGAAT 2874

QY 2424 CTTTCAGATATAGACAAATTCGTTTCTGTTTAAAGAAATGGTTATCCGGTGCAGTTCTATC 2483

Db 2875 CCAAATGATATTGATAAATTGATAAATTAAGGGTCTTGTCTCTGAGATCAATCCGGT 2934

QY 2484 ATACTGAAATTAAGGGGGCCCTTCTTCAAAATGTTTAGTGTGTGGTCACTCGCTCCGCTA 2543

|||

```
Db 2935 ATCCCTGATATGAAGGTAGCGTTTTCAAATCAACAGCTCTGCGATCA---TACAATGGCA 2991
Qy 2544 GTTACAGTTGTTAAAGGCGCGTTCAGAGAGCCAAACAAGGTGTGAAGAGCCAGAAATGTGCA 2603
Db 2992 GTGGAAATTTGATAGAGGATTTATACAGAGCCCGCTAGGTGTGAACGTTATGTTGTAAT 3051
Qy 2604 GCACGGAATGCTATGTCTTATTCACAAATCGATGCACTTTTGGCAAAATGAAGCATAGTG 2663
Db 3052 GAAACCAATTTCAATGTCATGATTCACAAACAGGTGTTCAATTCGCGAGATAAACAAGTCAT 3111
Qy 2664 CGTCTTCAAGAACTCCAGATGCCATCTCTGAAGGAGACCTCCACACACAGCTCAGCATG 2723
Db 3112 AAGTTACAGAACTCCAGACTTTGTGCTGTATGACAAACGCCCTCACTCTATCTCATTA 3171
Qy 2724 TGTATTATACAACTATGTTGATGCTGTGAAGCCCTGGAGATCGTATTGAGGTAAACAGGA 2783
Db 3172 TGTGTTTACGATGAATTAGTGATTTCTTCAGGGCGGGCGATCGTATTGAGGTGACTGCG 3231
Qy 2784 GTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTTACGAGCATTTGAT 2843
Db 3232 ACGTTCAAGTCCATCCCATTAGAGCTAATTCAGGCAACGGCTACTTAAAGTCGTTGTAT 3291
Qy 2844 AAGACTACATCGATTGCGTGCAGCTCAGAGAGTCTGACAGGGTCTGACGCAAACTGAA 2903
Db 3292 AAAACATACGTGTGTGTCACGTTAAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT 3351
Qy 2904 GATCCTATGGAGATGATAGAGAAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963
Db 3352 ACTTCTACTATTGACNAGAAATTAATGACAGACAGGTAGATCATACAGGTCGAGAA 3411
Qy 2964 TCAGAAAGTGTCTTAAGAACAAAGATTCAAAACCTTAAAGAGCTGTCCAAGTCCCGGCG 3023
Db 3412 GTAAGACAAATTTACTGATCAGGATTTAGCAAAAATTCGCGAGGTTCGCGCAAGAGAAT 3471
Qy 3024 ATTTATGATAGACTTTCAGGTGCTGCTGCCAAGCATTTGGAGCTTCAAGTATTAA 3083
Db 3472 TTATACAGTTTATTAGCCGCTCTATTGCCCCAAGTATTTCAGCGTAGAAGATGTCAG 3531
Qy 3084 AAGGGTCTCTTTGCGAGCTCTTTGGTGGGAAGGCTTAAGAAATTCATCTCGAGCATCT 3143
Db 3532 AAGGGTATATTACTTACGATATTGGCGGACGAAATAAACCTTTACCAAGGGTGGCCGT 3591
Qy 3144 TTCCGAGGTGACATCAATGTTTACTTGTGTTGGGACCCCTGGTACCAAGTAAATCTCAGCTG 3203
Db 3592 TATAGAGGTGACATAAATATTTACTTGTGGGGATCCTTCTACTTCCAAATCGCAAT 3651
Qy 3204 CTTGATATGTGCAAGAATAGTCTCTGTGGAACTCAGTGTGAGTGTGAGGAGGAGTTGCG 3263
Db 3652 TTGCAATACGTTTCAAAAATTTACTCTCGTGGTGTGTATCTTCCGGTAAAGGTTCACT 3711
Qy 3264 GCGGTTGGGCTGACAGCGTATGTAAACGAAGGATCCAGAACTCGAGAGCGGTATTGGAG 3323
Db 3712 GCGTGTGGTTTAACTGCTTATATTAACAGGGATGTCGACACAAACAATGTTGTTGGAA 3771
Qy 3324 AGCGGAGCTTTGGTCTTATGATGATCGTGGGATATGCTGTATCGATGAGTTTCGACAAAATG 3383
Db 3772 AGTGGTGCAATAGTATGTCGTAGTGGGTTGTTGTTGTTGACGAGTTTGATAAATG 3831
Qy 3384 TCTGATAATCCCGAAGCATGCTTTCATGAGGTAAATGAGGACAAACAACGGTATCTGTAGCC 3443
Db 3832 AGTGATTTCAAGGTCGCTTTCGACGAAGTCAATGGAACAGCAGACTATTTCAATCGCA 3891
Qy 3444 AAGGGGTATCATGCTCGTGAACGCTCGAGCTGTCTCTTCATGTCGATGCAATCT 3503
Db 3892 AAGCGGGAAATATCAACACTTAATGCGAAGTTCATTTTGGCCAGTGTAAACCCA 3951
Qy 3504 AGTGGTCCCGATACAAATGCGCCCTTCTGTGATTTGATAAACAATCCAGCTTCTCCAACT 3563
Db 3952 ATTGGTTCAGCTAGAACCCCAATTTGCTGTGACTGAAATATTGATCTACCGCCCCCA 4011
Qy 3564 CTACTTTCTAGATTTGATTTAATTTACTTAAGTCTCGACAAACAGCAGCAGCAAAACGAT 3623
Db 4012 CTACTTTGAGATTCGATCTGGTCTATCTTGTCTTGTGATAAGGTTGATGAGAAAAATGAC 4071
```

```
Qy 3624 CGTCTCTCGCCAGGCACTCTCGTGGCTTTTACACTAAGAAA-----CTATGAAGTTTCA 3677
Db 4072 AGAGAACTAGCCAAACACTTAACAATCTTTACCTGGAAGATAAGCCCGAAACATATTTCT 4131
Qy 3678 AAGCAGGAGCGCTTAGATCTACAACACTTACCGGCTATATCACCTATGCTCGTCAGCAT 3737
Db 4132 CAAGACGACGCTGTACCAAGTTGAATTTTAAAGATGATATATTAGTTATGCAAGGAGCAC 4191
Qy 3738 GTATCATCTCAATTAAGTGATGAAGCTGCTGAAGATTTTGATTAATTTGGCTATGTTGAGATG 3797
Db 4192 ATACACCCCAATATCAACGAGCGCTTAAGACTGAGCTGTTTCGTGCTTATGTAGGAATG 4251
Qy 3798 GCGCAAAAGGCAACTTTCTCGAAGCAGTATAAAGAGGTGATTAACAGCCACACCTCGGCAA 3857
Db 4252 AGAAAGATGGGTGACGATTCGAGATCCGATGGAAGAGAAATCACAGCTACCAACAGACAA 4311
Qy 3858 CTCGAAAGTATGATTTCTGATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGTA 3917
Db 4312 CTTGAAAGTATGATTTCTGTTGGCTGAGGCGCACGCAAAATGAAATTTGAAAAACGTCGTA 4371
Qy 3918 GAGAAAGTTGATGACGACAGAGCTGTGGCCCTTTTAGAGCTGCTTTTGACGCAATCTGCT 3977
Db 4372 GAGCTGGAGGATTTCAAGAGCCGTTAGATTAATTAGATCAGCCATAAAGATTATGCA 4431
Qy 3978 ACTGATCATGCAACAGGTACGATACATGATGGAATCTTATCAGACTGGAGTGTGGCCAGC 4037
Db 4432 ACAGACCCCTAAAACCGGTAAAATCGACATGAATTTAGTTTCAACACAGGTAATCAGTTAT 4491
Qy 4038 GAGCGTATTTCTGCGGCCCACTTG 4061
Db 4492 CAGAGAAAACCTACAGGAGGATTTG 4515
```

RESULT 10

SC9531X 27166 bp DNA linear PLN 18-APR-2005
S.cerevisiae chromosome XVI cosmid 9531.

LOCUS

SC9531X 27166 bp DNA linear PLN 18-APR-2005

DEFINITION

S.cerevisiae chromosome XVI cosmid 9531.

ACCESSION

Z49919.1 GI:887584

VERSION

CDC21 homologue; DS94; FlPo-ATP synthase g subunit; Grave's

KEYWORDS

disease; guanine-releasing protein; Mitochondrial carrier protein;

RNA polymerase I; RPA2; transfer RNA-Lys; zinc finger protein.

Saccharomyces cerevisiae (baker's yeast)

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 27166)

Bowman, S.

Unpublished

2 (bases 1 to 27166)

Barrell, B.G., Rajandream, M.A. and Walsh, S.V.

Direct Submission

Submitted (23-JUN-1995) Saccharomyces cerevisiae chromosome XVI

sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge

CB10 1RQ E-mail: barrell@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely

overlapped and those that are overlapped by more than 50%

of their length by a larger CDS have been omitted from this

analysis.

Details of the omitted CDS are available on request. The more

significant matches with motifs in the PROSITE database are

also included but some of these may be fortuitous. The length in

codons and the calculated codon adaptation index (CAI)

is given for each CDS.

Cosmid 9531 is overlapped at the 5' end by cosmid 9723, embl entry

SC9723, accession no. Z48951, and at the 3' end by cosmid 9367,

embl entry SC9367, accession no. Z49274.

FEATURES

source

1. .27166

/organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"

misc_feature	complement (11293..11361) /note="PS00028 Zinc finger, C2H2 type, domain"	
misc_feature	complement (11383..11445) /note="PS00028 Zinc finger, C2H2 type, domain"	
CDS	complement (113814..14143) /note="VP9531.07c unknown, len: 109, CAI: 0.08" /codon_start=1 /product="unknown" /protein_id="CAA90158.1" /db_xref="GI:887591" /translation="MSCTFSDLVGEYSCBSPTNVLICILBEEKAAPVVKDKP LFFILSSKENIFPCISRSSVSEVKSILRAKNFICPHAVIGFQKAKQLGRFPQLN GLLRLOH" complement (16908..17651) /note="VP9531.08c, unknown, len: 247, CAI: 0.11, putative zinc finger protein, contains 2 x PS00028 Zinc finger,"	
Query Match	10.8%; Score 470.8; DB 15; Length 27166;	
Best Local Similarity	55.6%; Pred. No. 1.5e-128;	
Matches	948; Conservative 0; Mismatches 747; Indels 9; Gaps 2;	
QY	2364 AAGCATATTAGCGCCAGACCTTCAATCTCAAGCATCGGTGCACATCGGTGAACCTCAAC	2423
Db	24269 AAGTTTATAAGTAAGGCCCTTACAATGTAGGTTCTTGTAAGGTATGCGCGAATTGAAT	24328
QY	2424 CTTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCATC	2483
Db	24329 CCAATGATATTGATTAATTGATAACTTTAAGGGTCTTGCTCGAGATCAACTCCGGTT	24388
QY	2484 ATACCTGAAATTAAAGGGGCGCTTCTTCAATGTTTATGTTGGTCACTCGGCTCCGCTA	2543
Db	24389 ATCCCTGATGAAGGTAGCGTTTTCAAATGCAACGCTCTCGCATCA---TACAATGGCA	24445
QY	2544 GTTACAGTTGTTAAAGGGCGGTTGAGGACCAACAGGTGTGAAAGCCAGCAATGTGCA	2603
Db	24446 GTGGAAATTGATAGAGGATTTACAAAGCCCGCTAGGTGGAACGTTGATTGATTAAAT	24505
QY	2604 GCACGAATGCTATGCTCTTATTCAATCGATGCACTTTTGCAAATTAAGCAGATAGTG	2663
Db	24506 GAACCCAAATCCATGCTGATTTCAACAGGTGTTCAATCGCAGATAACAGTCATT	24565
QY	2664 CGTCTTCAAGAAATCCAGATGCCATTCTCTGAAGGAGAGATCTCACACACAGTCAGCATG	2723
Db	24566 AAGTTACAGGAAATCCAGACTTTTGTGCTGATGGACAAAGCGCTCACTCTATCTCATTA	24625
QY	2724 TGTATTACAACTATGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAAACAGGA	2783
Db	24626 TGTGTTTACGATGAATTAGTGGATTCTTGCAGGGCGGCGCATCGTATTGAGGTGACTGGC	24685
QY	2784 GTTTTCAAGGCCATGCGAGTTCGAGTTGGTCCGAATCAACGAACATTTACGAGCATTTGAT	2843
Db	24686 ACGTTACGGTCCATCCCATTAGAGCTAATTCAGGCAACCGCTACTAAAGTCGTTGTTAT	24745
QY	2844 AAGACCTACATCGATTGCGTGCAAGTCTTGACAGGGGTGCACTGCAAACTGAA	2903
Db	24746 AAAACATACGTCGATGTTGGTCCAGGTATAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT	24805
QY	2904 GATCCTATGAGATGAGTAAGGAGAAATGATATGTTGCTGGGTATCATGAAAGTGATCT	2963
Db	24806 ACTTCTACTATTGAACAAGAAATTAATGCAGAACCAAGGTAGATCATACAGGTCGAAGAA	24865
QY	2964 TCAGAGCTGCTTAATGAGCAAGATTCAAAACTTTAAGAGCTGTCCAGCTCCCGGGC	3023
Db	24866 GTAAGAACAATTACTGATCAGGATTTAGCAAAAATTTGCGAGGTTGCGGCAAGAGAT	24925
QY	3024 ATTTATGATAGACTTTCAAGGTGCTGGCTCCCAAGCATTTGGGAGCTTGAAGATATTAA	3083
Db	24926 TTATACAGTTTATAGCCGCTCTATTGCCCCAAGTATTATACAGCTAGAGATGTCAG	24985
QY	3084 AAGGGTCTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAAGAAAATTCATCTGAGGATCT	3143
Db	24986 AAGGGTATATTACTTTCAGCTATTTTGGCGGCACGAATAAAACCTTTTACCAAGGGTGGCGGT	25045
QY	3144 TTCCAGGTGACATCAATGTTTATTCTGTTGGGACCCCTGGTACAGTAAATCTCAGCTG	3203
Db	25046 TATAGAGGTGACATATAATTTTACTTGTGGGATCTCTTACTTCCAAATCGAAAT	25105
QY	3204 CTTTCAGTATGTCACAGATAGCTCCTCGTGGAACTTACACATAGTGGGCGAGAACTTCG	3263
Db	25106 TTGCAATAGTTTCAAAAATTTACTCCTCGTGTGTATATCTTCGGGTAAAGGTTCACT	25165
QY	3264 CGGTTGGGCTGACAGCGGTATGTAAACAGAGATCCAGAACTCGAGAGACGGTATTGGAG	3323
Db	25166 GCGTTGGTTAACTGCTTATATTACAAGGATGTCGACACAAAACAACTTGTGTTGGAA	25225
QY	3324 ACGGAGCTTTGGTCTTATAGTATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG	3383
Db	25226 AGTGTGCACTTAGTATTGTTCTGATGGAGGTGTTGTTGTTATTTGACGAGTTGTATAAATG	25285
QY	3384 TCTGATAATGCCGGAAGCATGTTTCATGAGGTAAATGGAGCAACAAACGGTATCTGTAGCC	3443
Db	25286 AGTGAATCTTACAAGGTCCGTTTGACGAAGTCATGGAACAGAGACTATTTCAATCGCA	25345
QY	3444 AAAGGGGTATCATTTGCTCGCTGAACGCTCGGAGCTCTGTCTTGTGCATGTGCAATCTCT	3503
Db	25346 AAAGCGGGAATTTATCAACACTTAAATGCCAGAGTTCTATTTTGGCCAGTGTCAACCCA	25405
QY	3504 AGTGGTCCCGATACAATGCGCGCTTCTGTGATGTGATPAAACATCCAGCTTCTCCAACT	3563
Db	25406 ATTGGTTACGCTACAAACCCCAATTTGCTGTGACTGAAAAATATTGATCTACCGCCCCA	25465
QY	3564 CTACTTTCTAGATTTGATTTAATTTTAAATGCTCGACAAACACAGACGAGCAAAACGAT	3623
Db	25466 CTACTTTGAGATTCGATCTGCTCTATCTTGTCTTGTATGATGAGGTTGATGAGAAAAATGAC	25525
QY	3624 CGTGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAA-----CTATGAAGTTTCA	3677
Db	25526 AGAGAACTAGCCAAACACTTAACAAATCTTTACTTGAAGATAAGCCCGAACATATTTCT	25585
QY	3678 AAGCAGGACGCTTATAGATCTACAAACATTTACCGGTATATCACCTATGCTGCTCAGCAT	3737
Db	25586 CAAGACGACGTCTACCACTTGAATTTTAAACGATGATATTAGTTATGCAAAAGGAGCAC	25645
QY	3738 GTACATCTTACATTAAGTGATGAAGTCTGGAAGTTTGAATTAATGGCTATGTTGAGATG	3797
Db	25646 ATACACCCCAATTAATCCAGAGCCGCTAAGACTGAGCTTGTGCTGTTTATGAGAAATG	25705
QY	3798 CCCCAAAAGGGCAACTTTCTCGAAGCAGTAAAGGTGATTAACAGCCACACCTCCGCGCA	3857
Db	25706 AGAAGATGGGTGACGATTCGAGATCCGATGAGAGAGATCACAGCTACCAACAGACAA	25765
QY	3858 CTGAAAAGTATGATTGTTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA	3917
Db	25766 CTTGAAAAGTATGATTCTGTTGGCTGAGGCGCACGCCAAATGAAATTTGAAAAACGTCGTA	25825
QY	3918 GAGAAAGTTGATCGACAGACAGCTGTGCGCTTTTAGAGCTGCTGCTTTCAGGCAATCTGCT	3977
Db	25826 GAGCTGGAGGATGTTTCAAGAGCCGTTAGATTTAAATTTAGATCAGCCATAAAGATTATGCA	25885
QY	3978 ACTGATCATGCAACAGGTACGATGATGATGATCTTTATCACGACTGAGGTGTCGGCCAGC	4037
Db	25886 ACAGACCCCTAAACCCGGTAAAAATTCGACATGAATTTAGTTCCAAACAGGTAATCAGTTATT	25945
QY	4038 GAGCGTATTTCGTGGGCCCAACTTG 4061	
Db	25946 CAGAGAAAACCTACAGGAGGATTTG 25969	
RESULT 11		
SCCHRXVI		
LOCUS	SCCHRXVI 165536 bp DNA linear PLN 18-APR-2005	
DEFINITION	S.cerevisiae chromosome XVI 165536 bp sequence, cen rightwards.	
ACCESSION	271255	
VERSION	271255.1 GI:1279666	
KEYWORDS	ARO7; ATPase; beta-transducin; CCL1; CDC21 homologue; CDE1; CDE2;	

CDE3; centromere; CHL1; chorismate mutase; citrate synthase; coatomer zeta subunit; cyclin; delta element; DnaJ; DSS4; electron transport flavoprotein; eukaryotic initiation factor 5; F1Fo-ATP synthase g subunit; gamma adaptin; GLN1; glutamine synthetase; Grave's disease; guanine-releasing protein; Hauli; halotolerance; histidine RNA synthetase; HTS1; isocitrate lyase; LPT1; LTR; MAK3; MFS1; Mitochondrial carrier protein; N-acetyltransferase; NADPH-cytochrome P450 reductase; negative regulator; NHP6A; nonhistone chromosomal protein 6A; OSD1; phenylalanyl-tRNA synthetase; phosphotyrosine protein phosphatase; polyomitis; protamine P1; protein kinase; protein transport; regulatory protein; ribosomal protein L37a; RNA polymerase I; ROX1; RPA2; SEC8; SMK1; spermidine synthase; TIF5; TKL1; transfer RNA-Gly; transfer RNA-Lys; transfer RNA-Phe; transfer RNA-Ser; transketolase; tropomyosin; ubiquitin-activating enzyme; vacuolar H(+)ATPase 54 kDa subunit; VMA13; YMC1; YME1; YTA11; zinc finger protein.

SOURCE
ORGANISM
Saccharomyces cerevisiae (baker's yeast)
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 165536)
Badcock,K., Bowman,S., Churcher,C.M., Pearson,D., Rajandream,M.A., Walsh,S.V. and Barrell,B.G.
TITLE
JOURNAL
Submitted (19-APR-1996) Saccharomyces cerevisiae chromosome IV sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ. E-mail: barrell@sanger.ac.uk
COMMENT
SC8132X 248483 1 to 25540
SC9723 248951 100 to 17660
SC9531X 249919 104 to 27166
SC9367 249274 10 to 42190
SC3085 268111 790 to 10339 (PCR product)
SC9499X 249219 134 to 43776

The most significant matches using FASTA -o to a non-redundant database compiled from Swissprot 32 and PIR 46 and Wormpep9 are included.
All CDS over 100 codons have been analysed. CDS that are largely or completely overlapped by a larger CDS have been omitted from this analysis.
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons/amino-acids is given for each CDS as is the calculated codon adaptation index (CAI).
tRNA genes and LTRs are also included.

Notes:
This sequence has been compiled from the following submissions of individual contigs.

FEATURES
source
Location/Qualifiers
1..165536
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XVI"
<1..3032
/gene="YPL012W"
<1..3032
/gene="YPL012W"
/note="YPL012W, unknown, incomplete, len: 1009, CAI: 0.19"
/codon_start=3
/product="unknown"
/protein_id="CAA95029.1"
/db_xref="GI:1314064"
/db_xref="GOA:Q7LH12"
/db_xref="InterPro:IPR000357"
/db_xref="UniProt/TREMBL:Q7LH12"
/translation="SNLSKQLKAKTKEDINAVMSRLRITSVTSQWPSQIEP
LCDVLGVKSEQLVNSASFCESFMFTWAETITSSGLAENKYLRLVDITFAKPS
NVDLTLSKWTAVIKGMSVYATHQPLKALRKIPGVFMCTYLASEPVIYQAQASQ
LITSLSESQDLYTPFSDKFNVDDEIIISQAKTIFDLFSIRYSGREILKIL
VAAFNKFYRSPNPFKSLKIVDTRVNEEQFMDLRNSELIVIGASISAGMPETILAE
APLNLNDSSRPPGRALLPLRLDRTYTKNLANLATFQNELAPYIKSPQSKFDKVPESIQ

LRVQTIVDQIWSLTPRFCELPMDLRESFTDEFASELSSLLSYSEVELRTTICHALKVL
AESNVSVAESSHNVLQRFPISEAQNIEYLSKTNLLAVLNVYVOTTPNARS
YLITDQYLKINTEKLEKTFENNVCGLKNMNEESGNVNEKKKPOLPATTDLLI
ICMITYLPVSYSALPMSFSLTVNSADALQKRAYRIITKLSKSGTAVAQFISDI
ENMVDSSSVQTSAAKARITAKITIVELPLDHLDFIVRTVAEVLISKDVNEKSR
TAPDTLCMRKMNPEGIIKFIQIPGYDPTTPQSSISSEFFKIIISAGLJGESQHMV
SSSITGVACLVFEPKNDGILMDIYDTIELYTSNREIVKSAIGFTKVCVLGLPE
ELMRPKYPELLLLKLRSHBHTGHFAKVHIIERLIRFGYDVIETANFPREDRLLT
NIRKMRNKRKDEEVTGYSDVAATKGSFMSAFDEAVGSDRENDNGSDENVAG
GKMNGAKQFIVESGDPDLDSQTILAHISSTRPKFNKNQRARFNDDRNFNDFSEG
KLWVGQPKFSTNVDDLSAVTSINAYLEAVKSGPVRGQNKLFKRNKGKSDSDFGD
DDGDEKSRLMRGVNGQNKIKGKPKFKSRKQL"
/complement (3265. .4326)
/gene="YPL011C"
complement (3265. .4326)
/gene="YPL011C"
/note="YPL011C, unknown, len: 353, CAI: 0.12"
/codon_start=1
/product="unknown"
/protein_id="CAA95030.1"
/db_xref="GI:1314065"
/db_xref="GOA:Q12297"
/db_xref="InterPro:IPR006565"
/db_xref="InterPro:IPR009072"
/db_xref="SGD:S000005932"
/db_xref="UniProt/Swiss-Prot:Q12297"
/translation="MTYNDFYFALLRISLQLLKAQFDRARPSLVDMTDLAKFL
SLASEVSSIAQACQDDITIALQDITALENLGIVRPTNVLVDVENSELSSKMG
KFKDCIYQTLTDARTALPTVLLQSEESKESPLSAIPDYLQNLQKAKQKLET
KNRKTIELDILNNGDDMIKLVIAQRINMIERASKESQNPALPHIAGYKSSIL
SRHHHTITNEDRMPSTPRDEADLTIQENPFTVKLPIMRKNRLNLTLSPEDE
ELSELGEVGNOKSQSNNNEESFKENKSLTESPHGDDRLDISMEQFDSNDVTKWAQ
EDMDSTQRTSLDYGYF"
4641..5210
/gene="YPL010W"
4641..5210
/gene="YPL010W"
/note="YPL010W, unknown, len: 189, CAI: 0.19, 39, 7#
identity in 156 aa overlap to coatomer zeta subunit
COP2 BOVIN P35604"
/codon_start=1
/product="unknown"
/protein_id="CAA95031.1"
/db_xref="GI:1314066"
/db_xref="GOA:P53600"
/db_xref="InterPro:IPR008084"
/db_xref="InterPro:IPR011012"
/db_xref="SGD:S000005931"
/db_xref="UniProt/Swiss-Prot:P53600"
/translation="MSSLSLYTQAVLIDQGERIYAKYQPPHRSDEGHOLFNSV
KKQFEKQLYRKHKKQSEILIFEDHLVLYKEYIDITIYLVASLEENEIVLQQGFSA
IRGALDILNNGMDKNIQENYDMVLAIADITDNGVILETDSNTIASRVSKPPTNEP
QMALDLKGFGLGAWGFAKSFQERLQQL"
complement (5443. .8559)
/gene="YPL009C"
complement (5443. .8559)
/gene="YPL009C"
/note="YPL009C, unknown, len: 1038, CAI: 0.19, contains
glutamic acid rich region"
/codon_start=1
/product="unknown"
/protein_id="CAA95032.1"
/db_xref="GI:1314067"
/db_xref="GOA:Q12532"
/db_xref="InterPro:IPR008532"
/db_xref="SGD:S000005930"
/db_xref="UniProt/TREMBL:Q12532"
/translation="MKQRISSALDILLARELKQDLGVRLSNINYNIADSSKQFLKEN
KPSKLNINVDGRLIYLETFSRPIPTPSPFVYKLRKHLKAKRLTALKQVDDRLIV
LQFADGHFYLVLEFPFSGNVLLENRRIMALQRVVLEHNKVGQIYEMFDESLEFTN
NESADEIEKRNKAEYTSVNEWIKAVQAKYESDITVIKQINIQIGREGAKKKYKVP
SIHKLLSKVPHLSDDLKSLKLVFNIDPSRCLNLEETLSLLELSTQLEYNQLL
TTTDKRGVILAKRNVNISEKTDLEFIYDTFHPFKYINGGDTDSSCIIEVEGPN

RTLDKPFSTIESKVALRIQONQSOAKKIDDAARAENDRKIQALLDVQELNERKGLHI
 IENAPLIEBELTAVQGLIDQMDWNTIEKLIKSEQKGNRAQLNLPLNKKONKISV
 KLDLSLSEVTSDDNESQNTTSSDSDESSEKRSKSTKMSKNEKINVTI
 DLGLSAYANATYFNIKTTSAQOKKVKRNKAMNLEVKIDQOLKKLXDSHVLK
 KIRTPYFPEKYSFVSISSGFLVMGKSPAEITDIQYSKIEBDDDIYMSNFSHVIKN
 PKTSVPENTLMQAGLGMSSSEASWKKISSPFWCFKQVNSKFDGSDNSILPFGAFR
 LKENDNHLPAQILVMGFLWKVKTSGNEDGDDDEEBEEREBEEREBEEREBE
 BEKEEERKEEQOQDDSDNEVGLGKSGSDSTKNNSEPHDNLEKDIEKCTISS
 LTDSDGNAKANDNSSTORIIDEPCVPIISLTISNVSNRKGRKGLKIKQKYADQDE
 TERLRLALGTLKGIKEQOORKEEIMKREVEDRNKNRKEKORLQALKPTKKEKAR
 VNYDKHSELKPSLDKGDVDDIIIPVPAWPAALLKYKYVKIOPQSAKTKTLTEILH
 YPKSRPLDGSNDNEMDPQEHMIEKGLKEODLVLLLCVKVTIAGQKSTKNGNS
 SKKGKKCR

gene

/gene="CHL1"

CDS

/note="YPL008W, CHL1 gene, len: 861 CAl: 0.11
 SW:CHL1 YEAST P22516, CHL1 protein, contains PS00690
 DEAH-box subfamily ATP-dependent helicases signature"

/codon_start=1

/product="Chlp"

/protein_id="CAA95033.1"

/db_xref="GI:1314068"

/db_xref="GOA:P22516"

/db_xref="InterPro:IPR001945"

/db_xref="InterPro:IPR002464"

/db_xref="InterPro:IPR006554"

/db_xref="InterPro:IPR006555"

/db_xref="InterPro:IPR010614"

/db_xref="SGD:S000005929"

/db_xref="UniProt/Swiss-Prot:P22516"

/translation="MDKSEYSETFHYPIYDIQVQLMETVVRVLSGKKIALLSPT
 GTGKLSLICATYALWMLNKADIFRMEITNEDDSENLSDDPEPDWIDYTRKSVL
 QSKVLLNDYKHLMEINTTSCQKMDLKHGRYKSVDPFLRRKRGARHLDVSL
 BQDPTPRQYSENNDSKSTGRGRIKSDQYKLSLNSQIITLLDKIDGKVLSDPN
 NGRDPVNTQNPVKIYASRTYSQGLFTSOLRLPSFPSSPRDKVPDEKVLPLASK
 KQLCINPKVMKWTLEAINDACADLRHSEKGCIFYQNTNENRCHPDTLALRMISEI
 QIDELVPLGKSGICFPYASREALPIAEVVTLPYQYLSRSTRSLQINLNSIIVT
 DRAHLINTINGISQISLEDLKNCHEGIVTFYFNKPSRLPNRNLKNSLLMT
 LIQFVKNPKKIQKVSQIPDNDMPGNSIDTLNHLKLYIKVSKIAKYIDTYNOALKEP
 ESSKNENPKETHKQSVSOPILLFKVSQFLYCLTNLTSGQPFEEKNYGSIKYMALLEP
 KPESILNOAKCVLWAGTWPMSEFLSNLLPEVSEDIITLSCNHVIPKENLQTVIT
 NQPELEFTEKMSPLVNNHLFPFVLDUSKAVPKGGIVAFVPSQYLAHVQCWKQ
 NDRFATLNVNRIFYEAQGDHDIISGYSDSVABGRSLLLAIVGKLSGKINFPDQDL
 RAVVMGLFPFNI FSGELIVKRLAAKIMKSGGTBEASRATKEFMEINCMKAVNQS
 VGRATRHANDYANIYLLDVRNPNRPFKLSRWQDSINSEHTTHQVSISSTRKFFSMR
 SLNSR"

misc_feature

10169..10198

Query Match	Best Local Similarity	Matches	Conservative	Score	470.8;	DB 15;	Length	165536;
	55.6%;			Pred. No. 2e-128;				
		948;	0;	Mismatches	747;	Indels	9;	Gaps
2364	AAGCATAATTGAGCCAGACCTTTCAATCTCAAGCATCGGTGACATCGGTGAACCTCAAC	2423						
Db	67267	AAATTTTATAAAGTAAGGCCTTCAATGATAGGTTCTTGTAAAGGTATGCGCGAATTTGAAT	67326					
2424	CTTTCAGATATAGACAAATGGTTCTGTGTAAAGGAATGGTTATCCGGTGCAGTTCTATC	2483						
Db	67327	CCAAATGATATTGATAAATGATAAACTTAAAGGGTCTTGTCTCGAGATCAACTCCGGTT	67386					
2484	ATACCTGMAATTAAGGGGCTTCTTCAATGTTTAGTGTGTGCTCACTCGCCTCGCTA	2543						
Db	67387	ATCCCTGATGATGAAGGTAGCGTTTTCCTTCAAAATGCAACGCTCTGCGATCA---TACATGGCA	67443					
2544	GTTACAGTTTGTAAAGGGCGGTTGAGGACCCCAAGGTGTGAAAGGCCAGAAATGTGCA	2603						
Db	67444	GTGGAATTTGATAGAGGAGTTATACAAGACCCGCTAGGTGTGAACGTTATTGTTGAAT	67503					
2604	GCAAGGAATGCTATGCTCTTATTTCAATTCAGATGCACTTTTGCAATAAGCAGATAGTG	2663						
Db	67504	GAACCCAAATTCCTGCTCACTGATTTCAACAGGGTGTTCATTCGAGATAAACAAGTCAAT	67563					

QY	2664	CGTCTTCAAGAAACTCCAGATGCCATTCTGGAAGAGAGACTCCACACACAGTTCAGCATG	2723
Db	67564	AAGTTACAGGAACTCCAGACTTTGTGCTGTATGACAAACGCCCTCACTCTCTCAATTA	67623
QY	2724	TGTTTATACAACTATGTTGATGCTGTGAAGCTCGAGATCGTATTGAGGTACACGA	2783
Db	67624	TGTGTTTACGATGAATTAAGTGGATTCTTTCAGGGCGGGCGATCGTATTGAGGTGATGCG	67683
QY	2784	GTTTTCAAGGCCCATGGAGTTCGAGTTCGATCCGAATCAACGAATTAATACGACATTTGAT	2843
Db	67684	ACGTTACGTCATCCCCCATTTAGACTAATTCAGGCAACGGGTACTAAAGTCGTTGAT	67743
QY	2844	AAGACTCATCGATTGGTGCCACGTCGAAGAGTCTGACAGGGGTGCACTGCAAACTGAA	2903
Db	67744	AAAAATACGTCGATGTGTCACGTTAAATAAAGTTTCAGATAAGAGGTTAGACGTCGAT	67803
QY	2904	GATCTATGAGATGATGAAGAGATGATATGATGTATGCTGGGTATCATGAAGTATGATCT	2963
Db	67804	ACTTCTACTATTGAAACAAGAAATTAATGAGCAACAAGGTAGATCATAAACGAGTTCGAAGA	67863
QY	2964	TCAGAGCTGCTTAATGAAGCAAGATTCAAAACCTTAAAGAGCTGTCCAAGTCCCGGCG	3023
Db	67864	GTAAACAAAATTAAGTACGATGAGATTTAGCAAAAATTCGCGAGTTGCGGCAAGAGAT	67923
QY	3024	ATTTATGATAGACTTTCAAGGTGCGCTGCTCCAAGCATTTGGGAGCTTCGAAGATATTA	3083
Db	67924	TTATACAGTTTATAGCCGCTCTATTTGCCCAAGTATTTACGAGTAGAAGATGCAAG	67983
QY	3084	AAGGGTCTTTTGTGCGAGCTCTTTGGTGGGAAGGTAAAGAAATTCATCTCGGAGCATCT	3143
Db	67984	AAGGGTATATTAATTTAGCTATTTTGGCGGACGAATAAACTTTACCAAGGTGCGCGT	68043
QY	3144	TTCCGAGGTGACATCAATGTTTACTTGTGGGGACCTGGTACCAAGTAAATCTCAGCTG	3203
Db	68044	TATAGAGGTGACATAAATATTTACTTTGTGGGGATCTCTTACTTCCAAATTCGCAAAAT	68103
QY	3204	CTTCAGTATGTCACAGATAGCTCTCTGCTGGAATCTACACATAGTGGGCGAGGAATTCG	3263
Db	68104	TTGCAATAGTTTCAAAAATTAATCTCTGCTGCTGTATGATCTTCGGGTAAAGGTTCACT	68163
QY	3264	CGGGTTGGCTGACACGCTATGTAACGAGGATCCAGAACTCCGAGAGACGGTATTGGAG	3323
Db	68164	GGCGTTGGTTAACTGCTTATATTAACAGGGATGTCGACACAAACAACCTTGTTTTGGAA	68223
QY	3324	ACGCGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG	3383
Db	68224	AGTGGTGCAATAGTATTGCTGATGGAGGTGTTGTTGTTGATTTGACGAGTTTGATAAATG	68283
QY	3384	TCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTGCC	3443
Db	68284	AGTGAATCTTACAGGTGCGTCTTTCGACGAAAGTCATGGACAGCAGACTATTTCAATCGCA	68343
QY	3444	AAAGGGGTATCATTTGCCCTGCGTGAACGCTCGAGCTCTGCTCTGTCATGTGCAAACTCT	3503
Db	68344	AAAGCGGGAATTTATCAACAACCTTAATGCCAGAGTTCTATTTTGGCCAGTGTCTAACCCA	68403
QY	3504	AGTGGGTCCCGATACAAATGCGCGCCCTTCTGTGATTTGATAAACAATCCAGCTTCCTCAACT	3563
Db	68404	ATTGGTTACGCTACAAACCCCAATTTGCTGTGATGTAAGAAATATTGATCTACCGCCCCCA	68463
QY	3564	CTACTTTTCTAGATTTGATTTAAATTTACTTTAAATGCTCGACAAACACGACGAGCAAAATG	3623
Db	68464	CTACTTTTCTAGATTTGATTTAAATTTACTTTAAATGCTCGACAAACACGACGAGCAAAATG	68523
QY	3624	CGTGTCTCGCCAGGATCTCGTGGCTTTTACACTATGAAA-----CTATGAGTTTCA	3677
Db	68524	AGAGAACTAGCCAAAACACTTAAACAAATCTTTTACCTGGAAGATAAGCCCAACATATTTCT	68583
QY	3678	AAGCAGGACGCTTAGATCTTACAAAACACTTACCGCTATATCACCTTATGCTCGTCAGCAT	3737
Db	68584	CAAGACGACGCTGCTACCAAGTTGAATTTTAAACGATGATATATTAGTTATGCAAGGAGCAC	68643

Qy	3738	GTACATCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG	3797
Db	68644	ATACACCCCAATATACCCGAGCCGCTAAGACTGAGCTGTTGCTGCTTATGTAGGAATG	68703
Qy	3798	CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAGAGGTGATAACAGCCACCTCGGCAA	3857
Db	68704	AGAAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAATCACAGCTACCACAAGACA	68763
Qy	3858	CTCGAAGTATGATTCGTATCAGTCAAGCCCTAGCTCGAATCAGATTTTCTGAGTGGTA	3917
Db	68764	CTTGAAGTATGATTCGTGTTGGCTGAGGCGCAGCCAAATGAAATGAAAACGTCGTA	68823
Qy	3918	GAGAAAGTTGATGACAGCAAGCTGTGGCCCTTTTAGACGTCGCTTTGACGCAATCTGCT	3977
Db	68824	GAGCTGGAGGATGTTCAAGAGCCGTTAGATTAATTAGATCAGCCATAAAGATTTATGCA	68883
Qy	3978	ACTGATCATGCAACAGTACGATAGACATGATCTTATCAGACCTGGAGTGTGCGCCAGC	4037
Db	68884	ACAGACCCCTAAAACCGGTAAAATCAGCATGAATTTAGTTCAAAACAGGTAAATCAGTTATT	68943
Qy	4038	GAGCGTATTCGTGCGGCCCACTTG 4061	
Db	68944	CAGAAAAACTACAGGAGATTG 68967	
RESULT 12			
CR380957_02/c			
WPCOMMENT			
Sequence split into 13 fragments LOCUS CR380957 Accession CR380957			
	Fragment Name	Begin	End
	CR380957_00	1	110000
	CR380957_01	100001	210000
	CR380957_02	200001	310000
	CR380957_03	300001	410000
	CR380957_04	400001	510000
	CR380957_05	500001	610000
	CR380957_06	600001	710000
	CR380957_07	700001	810000
	CR380957_08	800001	910000
	CR380957_09	900001	1010000
	CR380957_10	1000001	1110000
	CR380957_11	1100001	1210000
	CR380957_12	1200001	1302002
Continuation (3 of 13) of CR380957 from base 200001 (CR380957 Candida glabrata strain CE			
Query Match 10.7%; Score 464.4; DB 15; Length 110000;			
Best Local Similarity 55.6%; Pred. No. 1.5e-126;			
Matches 938; Conservative 0; Mismatches 741; Indels 9; Gaps 2;			
Qy	2363	GAAGCATATTGAGCCAGACCTTTCAATCTCAAGCATCGGTGACATCGCTGAACTCAA	2422
Db	20935	GAAAGTTTACAAAGTAAGACCTTCAATGTTGAAACTCAAAAGGTATGAGAGAAATTGAA	20876
Qy	2423	CCCTTCAGATATAGACAAATGGTTCTCTGTTAAAGGAATGGTTATCCGGTGACGTTCTAT	2482
Db	20875	TCCAATGATATAGATAGTTGATTAAGCTGAAAGGGTTGGTTCTTAGCGGACACCTGT	20816
Qy	2483	CATACCTGAAATTAAGGGGCGCTTCTTCAATGTTTAGTGTGTGCTCACTCGCCCTCGCT	2542
Db	20815	TATACCAGACATGAAAGTAGCTTTTTCAAATGTAATTTTGTGATCAC---ACCATGGC	20759
Qy	2543	AGTTACAGTTGTTAAAGGGCGGTTGAGGAGCCAAAGAGGTGTGAAAAGCCAGATGTGC	2602
Db	20758	TGTTGAAATCGATCGTGTGTTAAATTCAGGAACCCAGCTAGATGTGAAACGTTGATTGTA	20699
Qy	2603	AGCAGGAATGCTATGCTCTTATTACAAATCGATGACCTTTTGCAAAATAAGCAGATAGT	2662
Db	20698	TGAGCTTAATCTATGAGCTTAATTTCAATATAGATGCTCTTTCGCTGATTAACAAGTTAT	20639
Qy	2663	GGCTTTCAAGAACTCCAGATGCAATTCCTGAAGAGAGACTCCACACAGATCAGCAT	2722
Db	20638	TAAACTGCAAGAACTCTCTGATTTGGTTCCAGATGGAACAAACACCTCACTCTGTTTCTCT	20579

Qy	2723	GTGTTTATACAACTATGGTTGATGCTGAAGCCTGAGATCGTATTTAGGTAAACAGG	2782
Db	20578	ATGTGTTTACGACGAGTTAGTTGATGTTGCGGTGCTGGTACAGAAATTGAATTACAGG	20519
Qy	2783	AGTTTTCAGGCCATGCGAGTTCCGAGTTGGTCCGAAATCAACGAAACATTTACGAGCAATGTA	2842
Db	20518	TACCTTCAGATCAATCCCATCAAGGCCAACTCTAGACAAGGTGCTCTTAAATCTTTATA	20459
Qy	2843	TAAAGCTTACATCANTTGGCTGCGAGCTCAAGAACTCTGACAGGGGTGACCTGCAAACTGA	2902
Db	20458	CAAAACATACATGATGTCGTCATGTAAAAAAGGTCTCGAACACTAGAAATTGGTGTGA	20399
Qy	2903	AGATCCTATGAGATGATGAAGAGGAATGATATGATGCTGGGTATCATGAAAAGTGATAC	2962
Db	20398	TGTTTCAACTATCGAACAGAACTATTGCAAAAATAAATACTAGATAATATGATGTGAGA	20339
Qy	2963	TTCAAGAGCTGCTAATGAAGCAAAAGATTCAAAAACCTTAAAGAGCTGTCCAAGCTCCCGG	3022
Db	20338	AGTAAGACAAATCTCTGATGCTGAAATCGAGAAGATAAAGCAGGTGTCACAAAGACCGA	20279
Qy	3023	CATTTATGATAGACTTTTCAAGGTGCTGGCTCCAAAGCATTTTGGAGCTTTGAAGATATTAA	3082
Db	20278	TCTTTATGATCTTCTTGGAGATCTATAGCGCCAAAGTATTTTATGAATTAGATGAGTCAA	20219
Qy	3083	AAAGGGTCTTCTTGGCCAGCTCTTTTGGTGGGAAGCTAAGAAAAATTCATCTGGAGCATC	3142
Db	20218	AAAGGATATATTACTACAAATTTTGGTGGTACTTAAACAAAACGTTTAAAGAGGGTGGTCG	20159
Qy	3143	TTTCCGAGGTGACATCAATGTTTACTTTTGGGGACCCCTGGTACCAGTAAATCTCAGCT	3202
Db	20158	TTATAGAGGTGATATTAAACATTTTGTGATGGTGACCCCTCCACATCCAAATCCAGAT	20099
Qy	3203	GCCTCAGTATGTGCAAGATAGCTCTCGTGGGAATCTACACTAGTGGCGGAGGAAGTTC	3262
Db	20098	CTTCAATATATTTTCAATAAATTTGCTCTAGAGGTGTCTACACATCTGGTAAGGGTTTATC	20039
Qy	3263	GGCGTTGGGCTGACAGGTATGTAACGAAGATCCAGAACTCCAGAGACGCTGATTTGA	3322
Db	20038	AGCAGTAGTCTAACGGCATATGTCACAAAGATGTCGATAGTAAACAAATTAGTTTATGA	19979
Qy	3323	GAGCGAGCTTTGGTTCTTAGTGTATCGTGGGATATGCTGTATCGATGAGTTTCGACAAAAT	3382
Db	19978	GAGTGTGCTTTAGTGTCTTCTGATGGTGGTATCTGTTGTATGATGAATTCGATAAAT	19919
Qy	3383	GTCTGATAATCCCGAAGCATGCTTTCATGAGGTATGAGAGCAACAAACGGTATCTGTAGC	3442
Db	19918	GAGTGAATCAACTAGATCTGTTCTTTCAGAAAGTATGAGGAACAAACAAACGATCTCCGTTGC	19859
Qy	3443	CAAAAGGGGTATCATTTGCTCGCTGAACGCTCGGACGCTCTGCTTGCATGTGCAAAATCC	3502
Db	19858	TAAAGCTGGTATTTACAAATGATGCTAGAAAGTTCCATTTTAGCTAGTGCACAAACC	19799
Qy	3503	TAGTGGGTCCCGATACAAATGCGCGCTTTCTGTGATTTGATTAACATCCAGCTTCTTCCAAAC	3562
Db	19798	TATTGGATCCCGTTTAAACCTTAATCTACCTGTTACTGAAATAATTGATCTTCCACCCC	19739
Qy	3563	TCTACTTCTTAGATTTGATTTAAATTTACTTAATGCTCGAACAACAGAGGAGGAGAAACGA	3622
Db	19738	ATTACTGTCAGAAATTCGACCTTGTTTATTTATTTTGGATAAAGTTGATGAAAGCACTGA	19679
Qy	3623	TCGTCGCTCGCCAGGATCTCGTGGCTTTTACACTATGAAAACATA-----TGAAGTTTC	3676
Db	19678	CCGTGATCTGCAAAACACTTGAACGCTTTTATCTAGAGGACAAAGCTGCAATGTTTAC	19619
Qy	3677	AAAGCAGGACCCCTTAGATCTTACAAACACTTACCGCGTATATCACTATGCTCGTCAGCA	3736
Db	19618	TACCGATGATGTTCTTCCCAATCGATTTCTTAAACAATACATCAATTTATGTCACACAGA	19559
Qy	3737	TGTACATCTCACAATTAAGTGTAGAGCTGTGGAAGATTTGATTAATGCTGTATGTTGAGAT	3796
Db	19558	TGTTTCATCCACTTGTAAACAGAGCAAGCAAAAATGAACTTGTAAAGGCTTATGTTGGCAT	19499
Qy	3797	CGGCAAAAAGGGCAACTTTCTCTGGAAGCAGTAAAAAGGTGATTAACAGCCACACCTCGGCA	3856

Qy	3542	TAACATCCAGCTTCCTCCAACTCTACCTTTCTAGATTGATTTAAATTTACTTTAATGCTCGA	3601	Db	54550	TGAAATTTGGCAATTTTCAAATGTAGCGTATGTGACCATACCCAGCGCGTAG---AAATTG	54606	
Db	44955	GAACATTTGATCTTCTCTACACTTTTGTCTCGTTTGTGATCTCGTCTATCTAATGCTTGA	45014	Qy	2554	TTAAAGGCGCGGTTTGAGGAGCCAAAGGTTGTGAAAGCCAGCAATGTGCGACACGGGAATG	2613	
Qy	3602	CAAAACAGACGAGAAAACGATGCTGCTCGCCAGGCACTCG-----TGGCTTTACA	3655	Db	54607	ATCGTGGTATTATCCAGGAGCCGTTGCGGTGCCCGCTGTTGCAATGTAATCAGAGAAACT	54666	
Db	45015	CCAACTCGATGAACAGATCGATCGTCTATCTGGCTACTCTATATCGCAAAACATGTACCTTGA	45074	Qy	2614	CTATGTCTCTTATTTCACAATCGATGCACCTTTTGCATAATTAAGCAGATAGTGCCTTCAAG	2673	
Qy	3656	CTATGAAAACATATGAAGTTTCAAGCAGGAGCGCTTAGATCTACAACACTTTACCGCGTA	3715	Db	54667	CCATGTGTTTAAATCCACATAGATGCTCTCTTTCAGATAGCAAGTTTATAAAGTTACDAG	54726	
Db	45075	CGACAAACCCGCCACTGCTGCCACTGGAGAGATCTTACCGGTGGAGTTTCTGACTCTGTA	45134	Qy	2674	AAACTCCAGATGCGCATTTCTGAAGGAGAGACTCCACACACAGTCAGCATGCTGTTTATACA	2733	
Qy	3716	TATCACCTATGCTCGTCAGCATGTACATCTCAATTAAGTGATGAAGCTGCTGAAGATTT	3775	Db	54727	AAACCCCGGATCTTGTCCAGATGGCCAGACTCCCACTCCGTCTCATTTGTATATATG	54786	
Db	45135	CCTGTCTGACGACGCCAATGCTTCCCTCAATCAACCCCGCAGCCCAAGACAGACT	45194	Qy	2734	ACACTATGTTGATGCTGTGAAGCCTGGAGATCGTATTTAGGTAAAGGAGTTTTCAGG	2793	
Qy	3776	GATTAATGGCTATGTTGAGATGGCCAAAAGGGCACTTTCTTGGAAAGCAGTAAAAAGT	3835	Db	54787	ATGAGTTGGTTTGAAGCTGTCAGGCGGGTGACAGATTTGAAGTGACAGGGATCTTCAGGT	54846	
Db	45195	TGTGGCGCATATGTGACATGCGAAAGCAGGAGGATGATCTCGTCTTCGGAACACG	45254	Qy	2794	CCATGGCAGTTTCGAGTTGGTCCGAATCAACGAACTTACGAGCATTTGATAAGACCTTACA	2853	
Qy	3836	GATTAACAGCACACCTCGGCACTCGAAAGTATGATTCGATCAGTGAAGCCCTAGCTCG	3895	Db	54847	CTATCCCATTAAGAGCAAAATCAGCGCAGCGTGCCTTAAATCCTGTACAGACCTACT	54906	
Db	45255	AATCACAGCCACCAACAGACAGCTGGAGAGTATGATTCGGCTGTGAGAGCATGCCAA	45314	Qy	2854	TGCAATGCGTGCACGCTCAAGAAAGTCTGACAGGGGTGCACTGCAAACTGAAGATCCTATGG	2913	
Qy	3896	AATGAGATTTTCTGAAGTGTGAGAAAGTTGATGACAGCAAGAGTGTGGCTCTTTAGA	3955	Db	54907	TGGATGTTGTTTCACTTCGAAAAGTTCAGCGCACGGCGCTCGATATTGA-----TACCT	54960	
Db	45315	GATCGCTTTCTGCTCCACAGCTCGAGGTTTCTGACGTGACGAAGCCGTGCGACTTATCAG	45374	Qy	2914	AGATGGATAAGGAGATGATATGATGCTGGGTATCATGAAGTGATATCTTCAGAAAGCTG	2973	
Qy	3956	CGTCGCTTTGCAAGCAATCTGCTACTGATCATGCAACAGTAGTACGATAGCATGGAATCTTAT	4015	Db	54961	CCACTGTTTGAACACGCAAAATCCTCCAAAACCAGATGGACAAATGTGGAGGAACTACGGAAG	55020	
Db	45375	AGCGCCATCAAGGATTACGCCACCGATCTGTTTCTTGGACGAATCGACATGATCTGAT	45434	Qy	2974	CTAATGAAGCAAAAGATTCAAAAACTTAAAGAGCTGTCCAACTCCCGGGCATTTTATGATA	3033	
Qy	4016	CACGACTGGAGTGTGCGGCGACGAGCGTATTCGTCGGGCCA	4056	Db	55021	TAACTGATGAAGTATCGCCAAGATCAATGCTGTTGACGACGCGCGGATGTTTACGAG	55080	
Db	45435	TCAGACAGGAACTTCTCTGCGCCAGCAACCGCGCAGGCCA	45475	Qy	3034	GACTTTTCAAGTCTGCTGGCTCCAAAGCATTTTGGGAGCTTTGAAGATATTTAAAAAGGCTCTTC	3093	
RESULT 14	AE016817_06	Sequence split into 15 fragments	LOCUS AE016817 Accession AE016817	Db	55081	TCTTGGCCAGATCCATTGGCGCAAGTATCTACGAACCTTGATGACATAAAGAAAGGAATTC	55140	
WPCOMMENT		Fragment Name	Begin	End	Qy	3094	TTTGGCAGCTCTTTGGTGGGAAGGCTAAGAAAAATTCATCTGGAGCATCTTTCCGAGGTG	3153
		AE016817_00	1	110000	Db	55141	TCCTGCAGCTATTCCGGCGGAGCAAAATAAAACATTTTACAAAGGGTGGCAGGTATAGGGGTG	55200
		AE016817_01	100001	210000	Qy	3154	ACATCAATGTTTTACTTGTTCGGGACCTGGTACAGTAAATCTCAGCTGCTTCAGATATG	3213
		AE016817_02	200001	310000	Db	55201	ATATAAATATATCTTTTGTGCGGTGATCCATCTACTTTCGAAGTCCCAATCTCTGCAGTACG	55260
		AE016817_03	300001	410000	Qy	3214	TGCAAGATAGCTCCTCGTGGAAATCTACACTAGTGGCGAGGAAGTTTCGGCGGTGGGC	3273
		AE016817_04	400001	510000	Db	55261	TCCCAAGATTGCTCTCTGCTGTTGTATACACATCTGGTAAAGGTTCTCTCTCGCGTTGGTC	55320
		AE016817_05	500001	610000	Qy	3274	TGACAGCGTATGTAAAGAGGATCCAGAACTCCAGAGACCGTATTTGGAGAGCGGAGCTT	3333
		AE016817_06	600001	710000	Db	55321	TAAACGCTATATTACAGAGATGCCGACCAAGCAATCCGTTGCGAGAGTGGCGCAC	55380
		AE016817_07	700001	810000	Qy	3334	TGGTCTTTAGTATGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGCTCATATG	3393
		AE016817_08	800001	910000	Db	55381	TAGTCCTATCTGACCGGCGGTGTGTCGATCGATGAAATTCGCAAGATGAGCGACTCGA	55440
		AE016817_09	900001	1010000	Qy	3394	CCCGAAGCATGCTTTCATGAGGTAATGGAGCAACAAAACGTTATCTGTAGCCAAAAGGGGTA	3453
		AE016817_10	1000001	1110000	Db	55441	CCCGTCTCTGTTTTCACAGATGATGATGAGCAGCAACCAATATCCGTGCCCAAGCGGGTA	55500
		AE016817_11	1100001	1210000	Qy	3454	TCATTGCTCTCGTGAACGCTCGAGCTGTGCTCTTTCGATGTGCAAAATCCTTAGTGGGTCCC	3513
		AE016817_12	1200001	1310000	Db	55501	TCATCACCACTCTAAATGCCAGAACGTCCTATTTGGCAAGTGGCAACCCCATAGGTTTCAC	55560
		AE016817_13	1300001	1410000	Qy	3514	GATACAAATGCGGCTTTCTGATTTGATACATCCAGCTTTCCTTCCAACTCTACTTCTTA	3573
		AE016817_14	1400001	1466886	Db	55561	GATACAAACCAATCTCCCGGTGACTGAGAACATTTGACTCTGCCCTCCATTTGCTGTCTA	55620
Continuation (7 of 15) of AE016817 from base 600001 (AE016817 Ashbya gossypii (= Eremoth				Qy	3574	GATTTGATTTTAAATTTACTTAAATGCTCGACAAACACAGACGACAAACCGATCGTCTGCTCG	3633	
Query Match			10.5%;	Score 457.8;	DB 15;	Length 110000;		
Best Local Similarity			54.9%;	Pred. No. 1.5e-124;				
Matches			975;	Conservative	0;	Mismatches 787;	Indels 15; Gaps 3;	
Qy	2374	AGGCAGACCTTCCAATCTCAAAGCATCGGTGCAATCGGTGAACCTCAACCCCTTCAGATA	2433	Db	54430	AGATCAGGCCGTACAAATTTGGATTTCAGAGGGGCGATGAGAGAGCTCAACCCCGAACGACA	54489	
Qy	2434	TAGACAAATGGTTTCTGTTTAAAGGAATGTTATCCGGTGCAGTTCTATCATACCTGAAA	2493	Db	54490	TCGATAAACCTAGTAAGCATAAAGGGTCTTGTTCTTCCTCCACTCTCTGTCATTTCCAGATA	54549	
Qy	2494	TTAAGGGGGCTTCTCTCAATGTTTGTAGTGTGGTGCATCGCTCGCGCTAGTTTACAGTTG	2553					

Db 55621 GATTGACTGTGATACCTGGTGTGGACAAGGCTCTCGAGTCCACCGATCGAGAAATTAG 55680
Qy 3634 CCAGGCACTCT-----CGTGGCTTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db 55681 CCAAGCACTGACGAGCTCTACCTAGAGGACAAGCCGCCCACTGTGTCGAGTGGACA 55740
Qy 3688 CTTTATGATCTCAAAACATTACCGGTATATCACTATGCTCGTCAAGATGATACATCTTA 3747
Db 55741 TCCCTCCGGTGCACCTTTCTCAAAATGATATCAAACTATGCTCAAGCAGCATATTCACCCGG 55800
Qy 3748 CATTAAAGTATGATGAGTGTGAAGATTTGATTAATGAGTGTGATGATGCGCCAAAGG 3807
Db 55801 TGATCAGAGAGGGGCAAGACGAGCTGTGCGGCCCTATGTGAACATGCGCAGCATGG 55860
Qy 3808 GCAACTTTCTCGAAGCAGTAAAGGTGATTAACAGCCACACCTCGGCAACTCGAAAGTA 3867
Db 55861 GCGAGCACTCGCGCAGATGAGAGAGAAATCACGCCCACTACTCGGAGCTGGAAGCA 55920
Qy 3868 TGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAAGTGTGAGAGAAAGTTG 3927
Db 55921 TGATCAGACTGTCCGAAGCGCACGCTAAGGTGCGCCCTTTCCGACGAGGTTGAGGTCAAG 55980
Qy 3928 ATGACGACAGAGCTGTGCGCCCTTTAGACGTCGCTTTGAGCAATCTGCTACTGATCATG 3987
Db 55981 ACGTCCAGAGCAGTGAGACTGATCAAGTCTGCCATCAAGGACTACGCAATAGACCCCA 56040
Qy 3988 CAACAGGTACGATAGACATGGATCTTATCAGCACTGGAGTGTGCGGACGAGCGTATTTC 4047
Db 56041 AGACAGGCAAGATCGACATGAACCTTATCCAAACCGGCAAGTCCGATGATCCAGCGCAGC 56100
Qy 4048 GTCCGCCCAACTTGTGTAGCTGTCTGCGAGAGCTTATAGCAGATAAAATTTCACTTGGCA 4107
Db 56101 TCCAGGAGGACCTTGCCCGTGAATTTGTCGCGCTGCTAACCGAAACGCTCTGCGGATGTCA 56160
Qy 4108 GCTCCTCTGGCTTGAAGACGAGTCACTCTTTGAGCA 4144
Db 56161 TCACCTACAATGAGCTAGCCAGATTGCTGAACGAGAA 56197

RESULT 15

CR382123 15/c

WPCOMMENT

Sequence split into 18 fragments LOCUS CR382123 Accession CR382123

Fragment Name	Begin	End
CR382123_00	1	110000
CR382123_01	100001	210000
CR382123_02	200001	310000
CR382123_03	300001	410000
CR382123_04	400001	510000
CR382123_05	500001	610000
CR382123_06	600001	710000
CR382123_07	700001	810000
CR382123_08	800001	910000
CR382123_09	900001	1010000
CR382123_10	1000001	1110000
CR382123_11	1100001	1210000
CR382123_12	1200001	1310000
CR382123_13	1300001	1410000
CR382123_14	1400001	1510000
CR382123_15	1500001	1610000
CR382123_16	1600001	1710000
CR382123_17	1700001	1753957

Continuation (16 of 18) of CR382123 from base 1500001 CR382123 Kluyveromyces lactis str.

Query Match 10.5%; Score 456; DB 15; Length 110000;
Best Local Similarity 55.8%; Pred. No. 5.1e-124;
Matches 936; Conservative 0; Mismatches 725; Indels 15; Gaps 3;
Qy 2363 GAAGCATATTGAGCGCAGCACTTTCAATCTCAAGCATCGGTGCATCGTGAATCTCAA 2422
Db 39609 GAAGTTGTTTCAAAGTGAGACCATATAATGTGGAACCAAAAGGGGTATGAGAGCTTAA 39550
Qy 2423 CCCTTCAGATATAGACAAATTTGTTCTTTAAAGGAATGGTTATCCGGTGCAGTTCTAT 2482

Db 39549 TCCTTAATGATATGTATAAATTGGTGAGTATCAAGGTTTAGTTCTAAGATCCACACCTAT 39490
Qy 2483 CATACCTGAAATTAAGGGGGCCCTTCTCAAATGTTTAGTGTGTGTCTACCTCCGCTCGCT 2542
Db 39489 TATCCAGATATAGTGTGCAATTTTCAAATGTAACGTATGCAATCA---TACTGTAGA 39433
Qy 2543 AGTTACAGTTGTTAAAGGGCGGTTGAGAGGCAACAAGGTGTGAAAGCCAGATGTGC 2602
Db 39432 AGTTGAGATCGATCGTGTGATTTATTTCAAGAAACAGTGAAGTGTCCAGAGTGTCTGTAA 39373
Qy 2603 AGCAGGAATGCTATGCTCTTTATTTCAAACTCGATGCACTTTTGCAAAATAAGCAGATAGT 2662
Db 39372 TAGCCCAAACCTCCATGGTTTTAGTTTCAATAGGTGCACTTTCCAAAGCAGCAAGTGA 39313
Qy 2663 GCGTCTTCAAGAAAATCCAGATGCCATTTCTGAAGAGAGACTCCACACACAGTCAAGCAT 2722
Db 39312 AAGTTTCAAGAAAACGCTGATCTAGTCCCTGATGGAACAACTCCACACTCTGTGTTCCCT 39253
Qy 2723 GTGTTTATACAACTATGTTGATGCTGTGAAGCCTGAGATCGTATTTGAGGTAAACAG 2782
Db 39252 ATGTGTATATGATGAGCTGGTTGATTCGTGCTGCTGATAGAAATTTGAGGTTTCAG 39193
Qy 2783 AGTTTTCAGGCCATGCGAGTTTCGAGTTGGTCCGATCAACGAACTTACGAGCATTTGA 2842
Db 39192 TATCTTTAGATCTATTCCAATCAGGTGCAATCCAAAGCAAGAGCCCTCAAAATCTCTATA 39133
Qy 2843 TAAAGACCTACATCGATGCGTGCAAGTCTCAAGAGTCTGACAGGGGTGCACTGCAAACTGA 2902
Db 39132 CAAGACATACATCGATGCTGTTTATATCCAAAGGTGGCTTAAGATAGATCGGTGGA 39073
Qy 2903 AGATCCTATGAGATGGAATAAGGAAATGATATGATGCTGGGTATCATGAAGTGAATAC 2962
Db 39072 CACTTCTACTGTGAAACAACAACTCTTGCAAGACCAAGATTGATAATGTGGAAGAGATT- 39014
Qy 2963 TTCAGAGCTGCTAATGAAGCAAGATTCAAAAATTAAAGAGCTGTCCAGCTCCGCG 3022
Db 39013 -----GAACTTTGAGTAGCAAGACATTAGGAGGATCAAAGAAACTGTCTCGCGATCGGA 38959
Qy 3023 CATTTATGATAGACTTTTCAAGGTCGTGCTCCAAGCATTTTGGGAGCTTTGAAGATATTAA 3082
Db 38958 TGTCTATGATGTTTTGTCACTGTTTCCATCGCACCAAGTATTTACGAGTTAGATGACGTCAA 38899
Qy 3083 AAAGGTCCTTTTCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTTCAATCTGGAGCATC 3142
Db 38898 GAAGGCAATCTTATTACAACTCTTTGGTGTGCTAATAAACTTTCAAGAAAAGGTGGAAG 38839
Qy 3143 TTTCCGAGTGACATCAATGTTTTTACTTTGGTGGGACCTGGTACCAAGTAATCTCAGCT 3202
Db 38838 ATATAGAGGTGACATTAATATTCTACTATGTGGTGATCCTTCGACCTCTAATAATCTCAA 38779
Qy 3203 GCTTCAGTATGTGCACAAGATAGCTCTCTGCGAATCTACACTAGTGGGCGAGGAAGTTC 3262
Db 38778 TTTACAATATGTTTCAATAGTTGCTCTCTGTTGTTTACACTTCGGGTAAAGGTTCCCT 38719
Qy 3263 GCGGTTGGGCTGCACAGCGTATGTAAAGAGGATCCAGAACTCCAGAGACGATGTTGGA 3322
Db 38718 TGTGTTGGTTTAACTGCTTATGTTTACAAGAGATGGGATACTAAGCAACTGGTGTGGA 38659
Qy 3323 GAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCCACAAAT 3382
Db 38658 AAGTGGTGTCTTAGTTCTCTCGACCGTGTGTGTTGTTATTTGATGAATTTGATAAGAT 38599
Qy 3383 GTCTGATATGCCGAGCATATGCTTCATGAGTAAATGGAGCAACAAACGATCTCTGAGC 3442
Db 38598 GAACGATATATCTAGATCAGTATATCATGAAGTTATGGAACCAACAACTCTATATGC 38539
Qy 3443 CAAAGGGGTATCAATGCGCTCGTGAACGCTCGAGCTGTGCTCTTGCATGTGCAATCC 3502
Db 38538 AAAAGCAGTATTTATGAACTTTAAATGCTAGAACTTCAATTTCTGGTGTAGTCCCAATCC 38479
Qy 3503 TAGTGGGTCCGATACAAATGCGCGCTTTCTGTGTATGATTAACATCAGCTTCTCTCAAC 3562

```

Db 38478 TATCAACTCTAGGTACAACTCTTAATCTTCCTGTGACTGAAAAATATTGATTTGCGCCTCC 38419
QY 3563 TCTACTTTCTAGATTGATTTAAATTTACTTAATGTCGACAAAACGACGAGCAAAACGA 3622
Db 38418 ATTATTATCAAGATTGATCTCGTTTATCTAGTGTCTTGACAAGGTCAACGAAGCGTCTGA 38359
QY 3623 TCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAAC-----TATGAAGTTTC 3676
Db 38358 TCGTGAACCTCGTAGCAATTTAACCCAGTTTATACCTGGAAGACAGACCGGATTCGTATC 38299
QY 3677 AAAGCAGGACGCTTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCA 3736
Db 38298 TCAAGGTGATATTCTACCCAGTTGAATTTCTTGACAGCCTACATCAACTATGCAAAACAAAA 38239
QY 3737 TGTACATCCTACATTAAGTGTAGAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGAT 3796
Db 38238 CATTCAACCCGGTAATAACTGAGTCAAGCAAGACTGAACCTTGTGAGGGCATATGTCGGTAT 38179
QY 3797 GCGCCAAAAGGGCAACTTTCTTGGGAAGCAGTAAAGGTGATAACAGCCACACCTCGGCA 3856
Db 38178 GAGAAGATGGGTGATGACAGTAGATCAGATGAAAAAAGAAATTCGCCAACAAACAGACA 38119
QY 3857 ACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGT 3916
Db 38118 GTTAGAAAGTATGATTAGACTTTTCAAGAGCTCATGCTAAGATGAGGTTAAGTGAAGAGT 38059
QY 3917 AGAGAAAGTTGATGACAGAGAGCTGTGGCCCTTTTAGACGTGCGCTTTTGCAGCAATCTGC 3976
Db 38058 TGAACCTTGAAGATGTGGAAGAGCAGTGAAGTGAATAAAGTCTGCCATCAAAAGATTACGC 37999
QY 3977 TACTGATCATGCAACAGGTACGATAGACATGGATCTTATCAGCACTGGAGTCTCGG 4032
Db 37998 TACGGATCCTAAACTGGTAAATTTGATATGAACCTTGTACAAACTGGTAAATCGG 37943

```

Search completed: December 6, 2005, 06:11:55
Job time : 14095 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 5, 2005, 23:36:15 ; Search time 1603 Seconds
(without alignments)
18077.407 Million cell updates/sec

Title: US-10-768-511-5
Perfect score: 4348
Sequence: 1 atggcgccgcactcacgt.....gcgaattccagagctgcg 4348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq 21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4348	100.0	4348	6	AAI67613	Cell cycl
2	4348	100.0	4348	10	AAD60813	Physcomit
3	880.2	20.2	2841	13	ADX53995	Plant ful
4	673	15.5	723	6	AAI67610	Cell cycl
5	673	15.5	723	10	AAD60810	Physcomit
6	634.2	14.6	1571	13	ADX36902	Plant ful
7	492.8	11.3	2970	10	ADB69927	C. neofo
8	490	11.3	1723	12	ADJ39580	Plant cDN
9	480.2	11.0	2733	6	ABZ32193	Candida a
10	470.8	10.8	2802	13	ADT47760	Bacterial
11	451.6	10.4	2625	8	ABT19388	Aspergill
12	451.6	10.4	3072	8	ABT21208	Aspergill
13	430.8	9.9	3273	6	ABL65258	Lung canc
14	430.8	9.9	3273	6	ABL65666	Lung canc
15	430.8	9.9	3273	6	ABL65259	Lung canc
16	430.8	9.9	3273	12	ADO19849	Human PRO
17	430.8	9.9	3273	12	ADO19847	Human PRO
18	430.8	9.9	3273	13	ADR26059	Breast ca
19	430.8	9.9	3273	13	ADU05807	Novel bro

20	430.4	9.9	3394	6	AAS94968	Human DNA
21	429.2	9.9	3395	10	ADJ56481	Adj56481 Frog cDN
22	427.6	9.8	2860	4	AAH17708	Human cDN
23	427.6	9.8	3248	13	ACN38801	Acn38801 Tumour-as
24	419	9.6	2842	3	AAF15678	Human pro
25	390.2	9.0	1105	13	ADX12413	Plant ful
26	352	8.1	2881	13	ADS49058	Bacterial
27	337.2	7.8	2944	8	ABT18794	Aspergill
28	337.2	7.8	3362	8	ABT20610	Aspergill
29	337.2	7.8	4944	8	ABT18200	Aspergill
30	337.2	7.8	5362	8	ABT20014	Aspergill
31	328	7.5	2692	4	ABL02965	Drosophil
32	321.4	7.4	595	13	ACN61030	Cotton gy
33	258.8	6.0	5265	12	ADP83382	Breast ap
34	255.4	5.9	3178	13	ACN41452	Human dia
35	243	5.6	770	4	AAF74193	DNA encod
36	228	5.2	1899	8	ABT18930	Aspergill
37	228	5.2	1913	8	ABT18336	Aspergill
38	228	5.2	2445	8	ABT20750	Aspergill
39	228	5.2	2608	8	ABT20152	Aspergill
40	228	5.2	3913	8	ABT17742	Aspergill
41	228	5.2	4608	8	ABT19556	Aspergill
42	226.6	5.2	1638	13	ADT47932	Bacterial
43	222.4	5.1	5065	4	ABL02964	Drosophil
44	220.8	5.1	2676	13	ADT48060	Bacterial
45	219.4	5.0	2280	13	ADT47813	Bacterial

ALIGNMENTS

RESULT 1
AAI67613
ID AAI67613 standard; cDNA; 4348 BP.
XX
AC AAI67613;
XX
DT 27-FEB-2002 (first entry)
XX
DE Cell cycle protein 2 (CC-2) encoding cDNA.
XX
KW Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;
KW CC-3; environmental stress; ss.
XX
OS Physcomitrella patens.
XX
PN WO200177354-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011294.
XX
PR 07-APR-2000; 2000US-0196001P.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Costa E SilvaO, Bohnert HJ, Van Thiel N, Chen R;
PI Sarria-Willan R;
DR WPI; 2002-049151/06.
DR P-PSDB; AAG66003.
XX
PT Novel Cell Cycle Stress-Related Protein useful for increasing tolerance
PT to environmental stress, is selected from Cell Cycle Proteins 1-3, or
PT their orthologs.
XX
PS Claim 4; Fig 2B; 90pp; English.
XX
CC The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),
CC isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1
CC protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSR and
CC encoding nucleic acids are useful for increasing tolerance to
CC environmental stress selected from salinity, drought and temperature, in

CC transgenic plants including monocot and dicot selected from maize, wheat,
CC rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,
CC canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,
CC tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,
CC tea, Salix species, oil palm, coconut, perennial grass and forage crops.
CC The nucleic acid is also useful for identifying organisms e.g. P. patens
CC in a mixed population of microorganisms. The nucleic acids are also
CC useful for evolutionary and protein structural studies. The proteins and
CC nucleic acids are useful as markers for specific regions of the genome.
CC The present sequence represents the cDNA encoding P. patens CC-2 protein
XX
SQ Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;

Query Match 100.0%; Score 4348; DB 6; Length 4348;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGCGCGCACCTCAGTGAAGGAATGACCTCTCTTGTCTGCGAGGTTCCATTCTT 60
DB 1 ATGGCGCGCGCACCTCAGTGAAGGAATGACCTCTCTTGTCTGCGAGGTTCCATTCTT 60
QY 61 TTTGGTTTTAGTTTGCAAACTTGATCGTGAAGTTGAGAAAAGGGCGGTTCTGTGTCT 120
DB 61 TTTGGTTTTAGTTTGCAAACTTGATCGTGAAGTTGAGAAAAGGGCGGTTCTGTGTCT 120
QY 121 TGAGGTGTTCTGTTGATGTTTCGTCATGGAATAATGATGCACTTGACATTTGAGCGG 180
DB 121 TGAGGTGTTCTGTTGATGTTTCGTCATGGAATAATGATGCACTTGACATTTGAGCGG 180
QY 181 TGTCTGCCCCATATCTTCGCAATCTGAAGAGTGTCTACGCCATTGCGCAAGTAACAT 240
DB 181 TGTCTGCCCCATATCTTCGCAATCTGAAGAGTGTCTACGCCATTGCGCAAGTAACAT 240
QY 241 CACCGAGTTTGAACAATCAGCCTCACCGTGCGCGGGCGAGGCCGTTACGGCAGACCC 300
DB 241 CACCGAGTTTGAACAATCAGCCTCACCGTGCGCGGGCGAGGCCGTTACGGCAGACCC 300
QY 301 CTACATCTGCAGTTTGAAGGAGAGGAGAGAAACGGATTCGCCCTGTGTAGGAGAGTC 360
DB 301 CTACATCTGCAGTTTGAAGGAGAGGAGAGAAACGGATTCGCCCTGTGTAGGAGAGTC 360
QY 361 GATCTCGCAGTTTAGGCAATCTGTTATAGTTTCCCTTACGATCGGGGACTCTCTGAA 420
DB 361 GATCTCGCAGTTTAGGCAATCTGTTATAGTTTCCCTTACGATCGGGGACTCTCTGAA 420
QY 421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCAGTCGGTACACCTATGGGTA 480
DB 421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCAGTCGGTACACCTATGGGTA 480
QY 481 CCCGATCGTTCCATCGTGGCAGCCACAGTACAAACAGGCGAGTGAGCTTGGTCCGAGG 540
DB 481 CCCGATCGTTCCATCGTGGCAGCCACAGTACAAACAGGCGAGTGAGCTTGGTCCGAGG 540
QY 541 GGAAGCCTCTACATCGGAGAGCTGCATCTCAATCCAGAGAACCCGGGATCGATCTCTTT 600
DB 541 GGAAGCCTCTACATCGGAGAGCTGCATCTCAATCCAGAGAACCCGGGATCGATCTCTTT 600
QY 601 CAAAGGAACTTAGTGTGATGGGCGTCCCTCTGAAATCTGCTGAGCCAGATGACCTTTGG 660
DB 601 CAAAGGAACTTAGTGTGATGGGCGTCCCTCTGAAATCTGCTGAGCCAGATGACCTTTGG 660
QY 661 GTGGAGAAATAGCTTATGTTTGGGGGACGAATGTTAAACATTCAGATGTCTTAGGGGA 720
DB 661 GTGGAGAAATAGCTTATGTTTGGGGGACGAATGTTAAACATTCAGATGTCTTAGGGGA 720
QY 721 TTCGTGCAATTTCTCCACAAATTAATGTTGAGTGTCTCATGATCTTAATCCAGATGATCC 780
DB 721 TTCGTGCAATTTCTCCACAAATTAATGTTGAGTGTCTCATGATCTTAATCCAGATGATCC 780
QY 781 AGATCATAGAGGAGACTGTGGAGCGGTGAGGAGGATATCTTAATATCGACATGTGAGACA 840
DB 781 AGATCATAGAGGAGACTGTGGAGCGGTGAGGAGGATATCTTAATATCGACATGTGAGACA 840

QY 841 TTTATGACCATGATCCTGATCTATACGCAAAAATTTGTTGATACCCACCTCGACATCATCC 900
DB 841 TTTATGACCATGATCCTGATCTATACGCAAAAATTTGTTGATACCCACCTCGACATCATCC 900
QY 901 CCCTGTTGGACACTGAGTGTGAGGAAGTTGCTACCTCTTTACTACCAAGTTTGAAGAC 960
DB 901 CCCTGTTGGACACTGAGTGTGAGGAAGTTGCTACCTCTTTACTACCAAGTTTGAAGAC 960
QY 961 ATATTGAGGCGAGACCTTTCAATCTCAAGGATCGGTGCACATGCGTGAACCTCAACCTT 1020
DB 961 ATATTGAGGCGAGACCTTTCAATCTCAAGGATCGGTGCACATGCGTGAACCTCAACCTT 1020
QY 1021 CAGATATAGACAAATTTGTTCTGTTAAAGGAATGTTATCCGTTGCGAGTCTTATCATAC 1080
DB 1021 CAGATATAGACAAATTTGTTCTGTTAAAGGAATGTTATCCGTTGCGAGTCTTATCATAC 1080
QY 1081 CTGAAATTAAGGGGCGCTTCTTCAAAATGTTTAGTGTGTGGTCACTCGCCTCGCTAGTTA 1140
DB 1081 CTGAAATTAAGGGGCGCTTCTTCAAAATGTTTAGTGTGTGGTCACTCGCCTCGCTAGTTA 1140
QY 1141 CAGTTGTTAAAGGGCGGTTGAGGAGCCAAACAGGTGTGAAAGCCAGAAATGTGAGCAC 1200
DB 1141 CAGTTGTTAAAGGGCGGTTGAGGAGCCAAACAGGTGTGAAAGCCAGAAATGTGAGCAC 1200
QY 1201 GGAATGCTATGCTCTTATTCACAAATCGATGCACTTTTGCAAAATGAGCAGATAGTCGTC 1260
DB 1201 GGAATGCTATGCTCTTATTCACAAATCGATGCACTTTTGCAAAATGAGCAGATAGTCGTC 1260
QY 1261 TTCAAGAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
DB 1261 TTCAAGAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
QY 1321 TATACAAACATATGTTGATGCTGTGAAGCCCTGTGAGATCGTATTTAGGTTAAACAGGTTT 1380
DB 1321 TATACAAACATATGTTGATGCTGTGAAGCCCTGTGAGATCGTATTTAGGTTAAACAGGTTT 1380
QY 1381 TCAGGCCATCGCAGTTGCTGAGTGTGCGCGCGCCACTCAGTGAAGTAATTCACCC 1440
DB 1381 TCAGGCCATCGCAGTTGCTGAGTGTGCGCGCGCCACTCAGTGAAGTAATTCACCC 1440
QY 1441 TCCTTGTCTCGCAGCGTTCCATCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1500
DB 1441 TCCTTGTCTCGCAGCGTTCCATCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1500
QY 1501 GTTGAGAAAAAGGGCGGTTGCTGTTGAGAGTGTCTTGTGATGTTGCTGATGAGAAA 1560
DB 1501 GTTGAGAAAAAGGGCGGTTGCTGTTGAGAGTGTCTTGTGATGTTGCTGATGAGAAA 1560
QY 1561 ATATATGATGCACTTGACATTCGAGCCGTTGCTGCTCCCATATCTTCCGCAATCTGAAGG 1620
DB 1561 ATATATGATGCACTTGACATTCGAGCCGTTGCTGCTCCCATATCTTCCGCAATCTGAAGG 1620
QY 1621 TGTCTACGCCATTGCGCAAGTAACATCAACGAGCTTCGACAAATGACAGCTCACCGCTGG 1680
DB 1621 TGTCTACGCCATTGCGCAAGTAACATCAACGAGCTTCGACAAATGACAGCTCACCGCTGG 1680
QY 1681 CCGGCGGAGGCGGCTACCGCAGACCCCTACATCTGCAAGTTCGAAAGGAGGAGAGAAA 1740
DB 1681 CCGGCGGAGGCGGCTACCGCAGACCCCTACATCTGCAAGTTCGAAAGGAGGAGAGAAA 1740
QY 1741 CCGATTCGCTCGTGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATCTGTTTATAGTT 1800
DB 1741 CCGATTCGCTCGTGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATCTGTTTATAGTT 1800
QY 1801 CCCCTTACGATCGGGGACTCTCTGAACTCTCTGGAATCTCCAGTGGCTACTCCGGTTTACG 1860
DB 1801 CCCCTTACGATCGGGGACTCTCTGAACTCTCTGGAATCTCCAGTGGCTACTCCGGTTTACG 1860
QY 1861 CTACCCAGTGGTACACTATGAGGTACCCCATCGTTCATCGTGGCAGCCACAGTACA 1920
DB 1861 CTACCCAGTGGTACACTATGAGGTACCCCATCGTTCATCGTGGCAGCCACAGTACA 1920
QY 1921 AACGCGCAGTGAGCTTGGTTCCCGAGGGAGCCCTCTATACCGGAGACGTCGATCTCAAT 1980

1921 AACAGCGAGTGGCTGGTCCAGGGAGGCTCTACATCGAGAGCGTCGATCTCAAT 1980
1981 CCAGAGAACCGGGGATCGATCTCTTCAAGGGACCTAGTCTGATGGGCGCTCCCTCTG 2040
1981 CCAGAGAACCGGGGATCGATCTCTTCAAGGGACCTAGTCTGATGGGCGCTCCCTCTG 2040
2041 AATCTGCTGAGCAGATGACACTTTGGGTGGAGATATCTTATGTTGGGGGAGCAATG 2100
2041 AATCTGCTGAGCAGATGACACTTTGGGTGGAGATATCTTATGTTGGGGGAGCAATG 2100
2101 TTAACATTCAGATGCTTTAGGGCGATTCGTCGATTTCTCCACAATATCGTTGAGTG 2160
2101 TTAACATTCAGATGCTTTAGGGCGATTCGTCGATTTCTCCACAATATCGTTGAGTG 2160
2161 CTCATGATCTTAATTCAGATGACATCATAGAGAGAGCTGTGGAGCGGTGAGGAGG 2220
2161 CTCATGATCTTAATTCAGATGACATCATAGAGAGAGCTGTGGAGCGGTGAGGAGG 2220
2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCTCTGATCTATACGCAAAA 2280
2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCTCTGATCTATACGCAAAA 2280
2281 TTGTTTCGATACCCACTCGACATCATCCCTCTGTTGGACACTGAGTGTCAAGAAAGTTGCTA 2340
2281 TTGTTTCGATACCCACTCGACATCATCCCTCTGTTGGACACTGAGTGTCAAGAAAGTTGCTA 2340
2341 CTTCTTTTATACCAACGTTTGAAGAGATATTAGAGCCAGACCTTTCAATCTCAAAAGCAT 2400
2341 CTTCTTTTATACCAACGTTTGAAGAGATATTAGAGCCAGACCTTTCAATCTCAAAAGCAT 2400
2401 CGGTGCACATGGGTGAACCTCAACCTTCAGATATAGCAAAATGGTTCTGTTAAAGGAA 2460
2401 CGGTGCACATGGGTGAACCTCAACCTTCAGATATAGCAAAATGGTTCTGTTAAAGGAA 2460
2461 TGGTTATCCGGTGCAGTCTTATCATCTGAAATTAAGGGGCGCTTCTTCAAAATGTTAG 2520
2461 TGGTTATCCGGTGCAGTCTTATCATCTGAAATTAAGGGGCGCTTCTTCAAAATGTTAG 2520
2521 TGTGTTGCTCACTCGCTCCGCTAGTTACAGTCTTAAAGGGCGGTTGAGAGCCAAACAA 2580
2521 TGTGTTGCTCACTCGCTCCGCTAGTTACAGTCTTAAAGGGCGGTTGAGAGCCAAACAA 2580
2581 GGTGTGAAAGCCAGAAATGTCAGCAGGAACTGCTATGCTTATTCACAAATCGATGCA 2640
2581 GGTGTGAAAGCCAGAAATGTCAGCAGGAACTGCTATGCTTATTCACAAATCGATGCA 2640
2641 CTTTTCGAAATAGCAGATAGTGGCTTCAAGAACTCCAGATGCCATTCCTGAAAGGAG 2700
2641 CTTTTCGAAATAGCAGATAGTGGCTTCAAGAACTCCAGATGCCATTCCTGAAAGGAG 2700
2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG 2760
2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG 2760
2761 GAGATCGTATGAGGTAAAGAGTTTCAAGGCATGCGCAGTTTGAAGTTGGTCCGAAATC 2820
2761 GAGATCGTATGAGGTAAAGAGTTTCAAGGCATGCGCAGTTTGAAGTTGGTCCGAAATC 2820
2821 AACGAACTTACGAGCATTTGATAAGACCTTACATCGATTCGATGCGTCAAGAAAGTCTG 2880
2821 AACGAACTTACGAGCATTTGATAAGACCTTACATCGATTCGATGCGTCAAGAAAGTCTG 2880
2881 ACAGGGTTCGACTGCAAACTGAACTCTTGGAGATGGATTAAGGAAATGATATGATG 2940
2881 ACAGGGTTCGACTGCAAACTGAACTCTTGGAGATGGATTAAGGAAATGATATGATG 2940
2941 CTGGGTATCATGAAAGTGATCTTCAAGAGCTGCTTAATGAAGCAAAAGATCAAAAATCTTA 3000
2941 CTGGGTATCATGAAAGTGATCTTCAAGAGCTGCTTAATGAAGCAAAAGATCAAAAATCTTA 3000
3001 AAGAGCTGTCCAGCTCCGGGCAATTTATGATAGACTTTCAAGTCCGCTGCTCCAGCA 3060

3001 AAGAGCTGTCCAGCTCCGGGCAATTTATGATAGACTTTCAAGTCCGCTGCTCCAGCA 3060
3061 TTTGGGAGCTTGAAGATATTAAGAGGTCTTTCTTTGCGAGCTCTTTGTTGGAGAGGCTA 3120
3061 TTTGGGAGCTTGAAGATATTAAGAGGTCTTTCTTTGCGAGCTCTTTGTTGGAGAGGCTA 3120
3121 AGAAAAATTCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTGGGGACC 3180
3121 AGAAAAATTCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTGGGGACC 3180
3181 CTGTTACCAAGTAATCTCAGCTGCTTCAAGTATGTCACAAGATAGTCTCTCTGTAATCT 3240
3181 CTGTTACCAAGTAATCTCAGCTGCTTCAAGTATGTCACAAGATAGTCTCTCTGTAATCT 3240
3241 ACACATAGTGGGCGAGGAAAGTTCCGGGCTGACAGCGTATGTAAACGAAGGATCCAG 3300
3241 ACACATAGTGGGCGAGGAAAGTTCCGGGCTGACAGCGTATGTAAACGAAGGATCCAG 3300
3301 AAACTCGAGAGACGGTATTTGGAGAGCGGAGCTTTGGTCTTAGTATGATCGTGGGATATGCT 3360
3301 AAACTCGAGAGACGGTATTTGGAGAGCGGAGCTTTGGTCTTAGTATGATCGTGGGATATGCT 3360
3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCAATGAGGTAATGG 3420
3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCAATGAGGTAATGG 3420
3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCAATTCGCTCGCTGAAACGCTCGACCGT 3480
3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCAATTCGCTCGCTGAAACGCTCGACCGT 3480
3481 CTGTCCTTCGATGTGCAAAATCTTAGTGGTCCCGATACAAATGCGCGCCCTTCTGTGATTTG 3540
3481 CTGTCCTTCGATGTGCAAAATCTTAGTGGTCCCGATACAAATGCGCGCCCTTCTGTGATTTG 3540
3541 ATAAATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTTAAATTAATGCTCG 3600
3541 ATAAATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTTAAATTAATGCTCG 3600
3601 ACAAACGACGAGCAAAACGATCGTCTGTCGCGAGGATCTCGTGGGCTTTACACTATG 3660
3601 ACAAACGACGAGCAAAACGATCGTCTGTCGCGAGGATCTCGTGGGCTTTACACTATG 3660
3661 AAAAATATGAAGTTTCAAGAGCAGGACGCTTAGATCTACAAACACTTACCGCGTATATCA 3720
3661 AAAAATATGAAGTTTCAAGAGCAGGACGCTTAGATCTACAAACACTTACCGCGTATATCA 3720
3721 CCTATGCTCGTCAGCATGTACATCTTACATTAAGTATGATGCTGCTGAAGATTTGATTA 3780
3721 CCTATGCTCGTCAGCATGTACATCTTACATTAAGTATGATGCTGCTGAAGATTTGATTA 3780
3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCTCGAAGCAGTAAAGAGTGTATAA 3840
3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCTCGAAGCAGTAAAGAGTGTATAA 3840
3841 CAGCCACACTCGGCAACTCGAAAGTATGATTCGATCAGTGAAGCCCTAGCTCGAATGA 3900
3841 CAGCCACACTCGGCAACTCGAAAGTATGATTCGATCAGTGAAGCCCTAGCTCGAATGA 3900
3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGACAGCAAGCTGTGCGCCTTTTGAAGCTCG 3960
3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGACAGCAAGCTGTGCGCCTTTTGAAGCTCG 3960
3961 CTTTTCGAGCAATCTGCTACTGATCATGCAACAGGTACGATAGCATGGATCTTATACGA 4020
3961 CTTTTCGAGCAATCTGCTACTGATCATGCAACAGGTACGATAGCATGGATCTTATACGA 4020
4021 CTGAGTGTGCGGCGAGCGGATTTGCTCGGGCCAACTTGTAGTGTCTCTGCGAGGC 4080
4021 CTGAGTGTGCGGCGAGCGGATTTGCTCGGGCCAACTTGTAGTGTCTCTGCGAGGC 4080
4081 TTATAGCAGATAAAATTTTCACTTGGCAGCTCTCTGCTTTGAAGACCAAGTCAGTCTTCTG 4140
4081 TTATAGCAGATAAAATTTTCACTTGGCAGCTCTCTGCTTTGAAGACCAAGTCAGTCTTCTG 4140

QY 4141 AGGATATCGGAGCCAAAGCAGTGTGGAGCTTAGTTGCGAGATATTAATAATGCTCTGG 4200
DB 4141 AGGATATCGGAGCCAAAGCAGTGTGGAGCTTAGTTGCGAGATATTAATAATGCTCTGG 4200
QY 4201 GTAGCCTCAAGGAGAGGCTTTCTTACTGTCTCATGTGACATAGTCAAGAGAGTTTGG 4260
DB 4201 GTAGCCTCAAGGAGAGGCTTTCTTACTGTCTCATGTGACATAGTCAAGAGAGTTTGG 4260
QY 4261 ACAGTTTCTAAGTGTTCGATCCATGAGTATTAATCTGACGAAAGGAAACCTCCAG 4320
DB 4261 ACAGTTTCTAAGTGTTCGATCCATGAGTATTAATCTGACGAAAGGAAACCTCCAG 4320
QY 4321 TTTTCCATGCGCAATTCCAGAGCTCGC 4348
DB 4321 TTTTCCATGCGCAATTCCAGAGCTCGC 4348

RESULT 2

AAD60813

ID AAD60813 standard; cDNA; 4348 BP.

AC AAD60813;

XX 15-JAN-2004 (first entry)

XX Physcomitrella patens cell cycle protein (CC-2) cDNA.

* KW Cell cycle stress-related protein; CCSRP; metal; cell cycle protein; CC;
KW transgenic plant; environmental stress; stress tolerance; salinity;
KW drought; temperature; chemical; oxidative stress; gene; ss.

* OS Physcomitrella patens.

XX Key Location/Qualifiers

FH CDS 147..4259

FT /*tag= a

* FT /product= "CC-2 protein"

FT /transl_except= (pos:1404..2816, aa:Gly-Pro)

XX US2003097675-A1.

XX 22-MAY-2003.

XX 06-APR-2001; 2001US-00828062.

XX 07-APR-2000; 2000US-0196001P.

XX (SILV/) COSTA E SILVA O.

PA (BOHN/) BOHNERT H J.

PA (THIE/) THIELEN N V.

PA (CHEN/) CHEN R.

PA (SARR/) SARRIA-MILLAN R.

XX Costa E SilvaO, Bohnert HJ, Thielen NV, Chen R, Sarria-Millan R;

XX WPI; 2003-765533/72.

DR P-PSDB; AA639980.

XX Novel cell cycle stress-related protein capable of conferring stress
PT tolerance such as tolerance towards salinity, drought, temperature,
PT chemical, pathogens, to plants upon over-expression.

XX Claim 16; Fig 2B; 62pp; English.

XX The invention relates to an isolated cell cycle stress-related protein
CC (CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
CC and its orthologues. Recombinant expression vector is useful for
CC producing a transgenic plant containing a CCSRP coding nucleic acid,
CC where expression of the nucleic acid in the plant results in increased
CC tolerance to environmental stress as compared to a wild type variety of
CC the plant which involves transforming a plant cell with the recombinant
CC expression vector, generating from the plant cell a transgenic plant with

CC an increased tolerance to environmental stress as compared to a wild type
CC variety of the plant. CCSRP is useful for conferring stress tolerance
CC such as tolerance towards salinity, drought, temperature, metal,
CC chemical, pathogens and oxidative stress or their combinations to plants.
XX The present sequence is Physcomitrella patens CC-2 cDNA

SQ Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;

Query Match 100.0%; Score 4348; DB 10; Length 4348;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCGCCGACCTCAGTGGAGAAATGACCTCTTGTGTCGAGCGTTCATTCCT 60

DB 1 ATGGCGCGCCGACCTCAGTGGAGAAATGACCTCTTGTGTCGAGCGTTCATTCCT 60

QY 61 TTTGGTTTTAGTTTGCATAATCTTGCATGCTGGAGTTGAGAAAAAGGGGGTTCGTGTCT 120

DB 61 TTTGGTTTTAGTTTGCATAATCTTGCATGCTGGAGTTGAGAAAAAGGGGGTTCGTGTCT 120

QY 121 TGAGGTGTTCTTGTGATGTTGTCATGGAATAATGATGCACTTGACATTTGGAGCGG 180

DB 121 TGAGGTGTTCTTGTGATGTTGTCATGGAATAATGATGCACTTGACATTTGGAGCGG 180

QY 181 TGTGTCGCCCATATCTTCCAAATCTGAAGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

DB 181 TGTGTCGCCCATATCTTCCAAATCTGAAGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

QY 241 CACGAGCTTCGACAAATGACGCTCACCCGTGGCGGGCGGAGGCGGTACGGCAGACCC 300

DB 241 CACGAGCTTCGACAAATGACGCTCACCCGTGGCGGGCGGAGGCGGTACGGCAGACCC 300

QY 301 CTACATCTGCAGTTCTGAAGAGAGGAGAGAAACGGATTTCGGCTGCTGCTAGGAGAGTC 360

DB 301 CTACATCTGCAGTTCTGAAGAGAGGAGAGAAACGGATTTCGGCTGCTGCTAGGAGAGTC 360

QY 361 GATCTCGCAGTTTAGGCAATTCGTTTATAGTTCCTCCCTACGATCGCGGGGACTCCTGGAA 420

DB 361 GATCTCGCAGTTTAGGCAATTCGTTTATAGTTCCTCCCTACGATCGCGGGGACTCCTGGAA 420

QY 421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTTACCCAGTCCGGTACACCTATGGGTA 480

DB 421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTTACCCAGTCCGGTACACCTATGGGTA 480

QY 481 CCCCATCTTCATCGCGAGAGTCCGATCTCAATCCAGAGAACCCGGGCGATCGATCTCCTT 600

DB 481 CCCCATCTTCATCGCGAGAGTCCGATCTCAATCCAGAGAACCCGGGCGATCGATCTCCTT 600

QY 601 CAAGGGAACCTAGTGTCTGATGGCGTCCCTCTGAAATCTGCTGAGCCAGATGACATTTGG 660

DB 601 CAAGGGAACCTAGTGTCTGATGGCGTCCCTCTGAAATCTGCTGAGCCAGATGACATTTGG 660

QY 661 GTGAGAAATATGCTTATGTTTGGGGGAGCAATGTTTAAACATTCCAGATGCTAGGCGCA 720

DB 661 GTGAGAAATATGCTTATGTTTGGGGGAGCAATGTTTAAACATTCCAGATGCTAGGCGCA 720

QY 721 TTGCTGAGTTTCTCCACAATTTATCGTTGAGTGTCTCATGATCTTAATTCAGATCATCC 780

DB 721 TTGCTGAGTTTCTCCACAATTTATCGTTGAGTGTCTCATGATCTTAATTCAGATCATCC 780

QY 781 AGATCATAGAGAGACTGTGGAGCGTGGAGGAGTACTCTAAATATCGACATGTGACACA 840

DB 781 AGATCATAGAGAGACTGTGGAGCGTGGAGGAGTACTCTAAATATCGACATGTGACACA 840

QY 841 TTTATGACCATGATCTCTGATCTATACGCAAAATTTGTTGATACCCACTCGACATCATCC 900

DB 841 TTTATGACCATGATCTCTGATCTATACGCAAAATTTGTTGATACCCACTCGACATCATCC 900

QY 901 CCTGTTGGACACTGAGTGTGAGGAGTGTGCTACCTCTTACTACCAAGCTTTGAGAGC 960

901	DB		960	DB	
961	QY		1020	QY	
961	DB		1020	DB	
1021	QY		1080	QY	
1021	DB		1080	DB	
1081	QY		1140	QY	
1081	DB		1140	DB	
1141	QY		1200	QY	
1141	DB		1200	DB	
1201	QY		1260	QY	
1201	DB		1260	DB	
1261	QY		1320	QY	
1261	DB		1320	DB	
1321	QY		1380	QY	
1321	DB		1380	DB	
1381	QY		1440	QY	
1381	DB		1440	DB	
1441	QY		1500	QY	
1441	DB		1500	DB	
1501	QY		1560	QY	
1501	DB		1560	DB	
1561	QY		1620	QY	
1561	DB		1620	DB	
1621	QY		1680	QY	
1621	DB		1680	DB	
1681	QY		1740	QY	
1681	DB		1740	DB	
1741	QY		1800	QY	
1741	DB		1800	DB	
1801	QY		1860	QY	
1801	DB		1860	DB	
1861	QY		1920	QY	
1861	DB		1920	DB	
1921	QY		1980	QY	
1921	DB		1980	DB	
1981	QY		2040	QY	
1981	DB		2040	DB	
2041	QY		2100	QY	
2041	DB		2100	DB	
2101	QY		2160	QY	
2101	DB		2160	DB	
2161	QY		2220	QY	
2161	DB		2220	DB	
2221	QY		2280	QY	
2221	DB		2280	DB	
2281	QY		2340	QY	
2281	DB		2340	DB	
2341	QY		2400	QY	
2341	DB		2400	DB	
2401	QY		2460	QY	
2401	DB		2460	DB	
2461	QY		2520	QY	
2461	DB		2520	DB	
2521	QY		2580	QY	
2521	DB		2580	DB	
2581	QY		2640	QY	
2581	DB		2640	DB	
2641	QY		2700	QY	
2641	DB		2700	DB	
2701	QY		2760	QY	
2701	DB		2760	DB	
2761	QY		2820	QY	
2761	DB		2820	DB	
2821	QY		2880	QY	
2821	DB		2880	DB	
2881	QY		2940	QY	
2881	DB		2940	DB	
2941	QY		3000	QY	
2941	DB		3000	DB	
3001	QY		3060	QY	
3001	DB		3060	DB	
3061	QY		3120	QY	
3061	DB		3120	DB	

QY	3121	AGAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTTACTTTGTTGGGACC	3180	QY	4201	GTAGCTCCAAAGGAAGGCTTTCTTACTGTCATGTTGACATAGTCAAGAGAGTTTGAG	4260
Db	3121	AGAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTTACTTTGTTGGGACC	3180	Db	4201	GTAGCTCCAAAGGAAGGCTTTCTTACTGTCATGTTGACATAGTCAAGAGAGTTTGAG	4260
QY	3181	CTGGTACCAAGTAAATCTCAGCTGCTTCTAGTGTGTCACAGATAGTCTCTCGTGGAACTC	3240	QY	4261	ACAGTTTCTAACTGTTGGAATCCATGAGCTATAACTCTGAACGAAAGGAAACCTCCAG	4320
Db	3181	CTGGTACCAAGTAAATCTCAGCTGCTTCTAGTGTGTCACAGATAGTCTCTCGTGGAACTC	3240	Db	4261	ACAGTTTCTAACTGTTGGAATCCATGAGCTATAACTCTGAACGAAAGGAAACCTCCAG	4320
QY	3241	ACACTAGTGGCGAGGAAGTTGCGGGTTGGGCTGACAGCGTATGTAACGAAGATCCAG	3300	QY	4321	TTTCCCATGCGCAATTTCCAGAGCTCGC	4348
Db	3241	ACACTAGTGGCGAGGAAGTTGCGGGTTGGGCTGACAGCGTATGTAACGAAGATCCAG	3300	Db	4321	TTTCCCATGCGCAATTTCCAGAGCTCGC	4348
QY	3301	AAATTCGAGAGCGGTATTGGAGAGCGAGCTTTGTTCTTAGTGATCGTGGGATGCT	3360	RESULT 3			
Db	3301	AAATTCGAGAGCGGTATTGGAGAGCGAGCTTTGTTCTTAGTGATCGTGGGATGCT	3360	ID	ADX53995 standard; cDNA; 2841 BP.		
QY	3361	GTATCGATGAGTTCGACAAAATGTCGTGATAATGCCGAGCATGCTTCATGAGGTAATGG	3420	XX	ADX53995;		
Db	3361	GTATCGATGAGTTCGACAAAATGTCGTGATAATGCCGAGCATGCTTCATGAGGTAATGG	3420	XX	AC		
QY	3421	AGCAACAAACGGTATCTGTAGCAAAAGGGGGTATCATTCGCTCGTGAACGCTCGGACGT	3480	XX	21-APR-2005 (first entry)		
Db	3421	AGCAACAAACGGTATCTGTAGCAAAAGGGGGTATCATTCGCTCGTGAACGCTCGGACGT	3480	XX	Plant full length insert polynucleotide seqid 28735.		
QY	3481	CTGTCTTCATGTGCAAAATCTAGTGGTCCGATACAAATGCGCGCTTCTGTGATG	3540	DE	plant protectant; plant growth regulant; gene therapy; plant;		
Db	3481	CTGTCTTCATGTGCAAAATCTAGTGGTCCGATACAAATGCGCGCTTCTGTGATG	3540	KW	recombinant DNA construct; physical array; plant breeding marker;		
QY	3541	ATAACATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAACTTAAATGCTCG	3600	KW	cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;		
Db	3541	ATAACATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAACTTAAATGCTCG	3600	KW	extreme osmotic condition; pathogen tolerance; pest tolerance;		
QY	3601	ACAAACGAGCAGCAAAAACGATCGTCTCTCCAGGCAATCTCGTGGCTTTTACATAG	3660	KW	growth rate; cell cycle pathway; disease resistance;		
Db	3601	ACAAACGAGCAGCAAAAACGATCGTCTCTCCAGGCAATCTCGTGGCTTTTACATAG	3660	KW	galactomannan production; lignin production; plant growth regulator;		
*QY	3661	AAACTATGAAGTTTCAAGCAGGAGCGCTTGAATCTACAAACACTTACCGGCTATATCA	3720	KW	yield; plant growth; plant development; seed oil; protein yield;		
Db	3661	AAACTATGAAGTTTCAAGCAGGAGCGCTTGAATCTACAAACACTTACCGGCTATATCA	3720	XX	protein content; gene; ss.		
QY	3721	CCTATGCTCGTCAGCATGTACATCTTACATTAAGTATGATGAAGCTGTGAAAGATTTGATTA	3780	OS	Unidentified.		
Db	3721	CCTATGCTCGTCAGCATGTACATCTTACATTAAGTATGATGAAGCTGTGAAAGATTTGATTA	3780	XX	US2004034888-A1.		
QY	3781	ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCTGGAAAGCAGTAAAAAGGTGATAA	3840	XX	19-FEB-2004.		
Db	3781	ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCTGGAAAGCAGTAAAAAGGTGATAA	3840	XX	28-APR-2003; 2003US-00425114.		
QY	3841	CAGCCACACTCGGCAACTCGAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA	3900	PF	06-MAY-1999; 99US-00304517.		
Db	3841	CAGCCACACTCGGCAACTCGAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA	3900	PR	05-NOV-2001; 2001US-00985678.		
QY	3901	GATTTTCTGAAGTGTAGAGAAAGTTGATGACAGCAAGCTGTGCGCCCTTTAGACGTG	3960	XX	(LIU/) LIU J.		
Db	3901	GATTTTCTGAAGTGTAGAGAAAGTTGATGACAGCAAGCTGTGCGCCCTTTAGACGTG	3960	PA	(ZHOU/) ZHOU Y.		
QY	3961	CTTTGAGCAATCTGCTACTGATCATGCAACAGGTACGATGAGTCTTATCACGA	4020	PA	(KOVA/) KOVALIC D K.		
Db	3961	CTTTGAGCAATCTGCTACTGATCATGCAACAGGTACGATGAGTCTTATCACGA	4020	PA	(SCRE/) SCRE S E.		
QY	4021	CTGGAGTGTGCGCCAGCAGCGATTTCTGTCGGGCCAACTTGTAGCTGTCTGCGAGAGC	4080	PA	(TAB/) TABASKA J E.		
Db	4021	CTGGAGTGTGCGCCAGCAGCGATTTCTGTCGGGCCAACTTGTAGCTGTCTGCGAGAGC	4080	PA	(CAO/) CAO Y.		
QY	4081	TTATAGCAGATAAAATTTTCCACCTGGCAGCTCTCTGGCTTGAAGACAGTCTGCTTCTG	4140	XX	Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;		
Db	4081	TTATAGCAGATAAAATTTTCCACCTGGCAGCTCTCTGGCTTGAAGACAGTCTGCTTCTG	4140	XX	WPI; 2004-180133/17.		
QY	4141	AGGATATCCGGACCAAGCAGTGTGAGCTGATTTGTCAGGATATTAATAATGCTCTGG	4200	XX	New recombinant DNA construct, useful for improving plant tolerance to		
Db	4141	AGGATATCCGGACCAAGCAGTGTGAGCTGATTTGTCAGGATATTAATAATGCTCTGG	4200	XX	cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or		
				XX	pests, for conferring increased resistance to plant disease, or for		
				XX	improving yield.		
				PS	Claim 1; SEQ ID NO 28735; 15pp; English.		
				CC	The invention describes a recombinant DNA construct comprising a		
				CC	polynucleotide consisting of a sequence encoding an amino acid sequence		
				CC	available in electronic form from the US patent office at		
				CC	ftp://seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide		
				CC	of the invention are also useful in physical arrays of molecules and as		
				CC	plant breeding markers. The recombinant DNA construct is useful for		
				CC	improving plant tolerance to cold, heat, drought, herbicides, extreme		
				CC	osmotic conditions, pathogens or pests, for manipulating growth rate in		
				CC	plant cells by modification of the cell cycle pathway, for conferring		
				CC	increased resistance to plant disease, for producing galactomannan,		
				CC	lignin or plant growth regulators, for increasing the rate of homologous		
				CC	recombination in plants, for improving yield by modification of		

CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.

Query Match	20.2%	Score 880.2;	DB 13;	Length 2841;
Best Local Similarity	65.3%	Pred. No. 7.4e-275;		
Matches 1334; Conservative	0;	Mismatches 688;	Indels 21;	Gaps 2;

Qy	2217	GAGGATACTCTAAATATTCGACATGTCAGACATTTATGACCATGATCTGTATCTATACGCA	2276
Db	625	GGGGAGTCGCTCGACGTCGATGCGCAGCAGCGTGTGTGACCGACCGACCTCTACAGC	684
Qy	2277	AAAATTGTCGTATACCCACTCGACATCATCCCCTGTTTGACACTGAGTGTCAAGGAATT	2336
Db	685	AAGATGTTTCGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCTGCTCATGAGCCTC	744
Qy	2337	GCTACCTCTTTACTTACCAAGCTTTTGAAAGCATATTGAGGCCAGACCTTTCAATCTCAAA	2396
Db	745	GTCCGGCGCATCGAGCCGCTCTTCGAGAGCAATCCAGACCGAGGATCTACACCTCAAG	804
Qy	2397	GCAATCGGTGCACATGCGGTGAACCAACCTTCAGATATAGACAAATTTGTTTCTGTAA	2456
Db	805	TCGTCCATTTGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAG	864
Qy	2457	GGAAATGTTATCCGGTCCAGTTCATCATACCTGAAATTAAGGGGCGCTTCTTCAAAATGT	2516
Db	865	GGTATGATAAATAGATCGAGTCGGTCAATACCGGAGCTCAAGGAGGCTGTGTTCCCGTGC	924
Qy	2517	TTAGTGTGTGTCACTCGCCTCCGCTAGTTTACAGTGTGTTAAAGCGCGGTGTGAGGAGCCA	2576
Db	925	CTGTTTGTGTGTTTCTACTCAGACCCGTCTATGTTGATAGAGGAAGTAACTGACCA	984
Qy	2577	ACAAGGTGTGAAAAGCCAGAAATGTCAGCACGGAATGCTATGTTCTTTATTCACAAATCGA	2636
Db	985	CACATTTGTTCAGAAAGAACAAATGTAAAGCCACAAATTTCTATGACCTTAGTGCAACAG	1044
Qy	2637	TGCACTTTTGCAAATTAAGCAGATAGTGGTCTTCAAGAAACTCCAGATGCCATTCCTGAA	2696
Db	1045	TGCAGATTTTCAGACAAGCAGATCATAAAGTTTGAGGAAACACACAGACGAGATACCGAA	1104
Qy	2697	GGAGAGACTCCACACACAGTCAGCATGTGTTTATACAAACTATGTGTTGATGCTGTCAAG	2756
Db	1105	GGTGGCACTCCACATACAGTTAGTGTCTTGATGCAATGATAGCTTGTGTGATGCTGGAA	1164
Qy	2757	CCTGGAGATCGTATTGAGGTAAACAGGAGTTTTTCAAGGCCATGCGAGTTCCAGTTGGTCCG	2816
Db	1165	CCTGGAGATAGGTTGAGATAACTGGAAATACAGAGCTATGATATTCGAATTGGACCA	1224
Qy	2817	AATCAACGAACATTTACGAGCATTTGTAAGAACCCTACATTCGATTCGGTGCACGTCACGAAG	2876
Db	1225	ACTCAAGGACAGTGAAGTCTATATTCAAGACATATATTGATTTCCTTCCATCAATAAGAG	1284
Qy	2877	TCTGCAGGGGTGCAGCTGCAAACTTGAGATCCTATGGAGATGGATATAGGGAATGATG	2936
Db	1285	ACAGACAAGTCTAGGCTTCATGTGGAGGACACCATGGATATTTGATAATCTTAACGCTAGC	1344
Qy	2937	TATGTCGGGTATCATGAAAGTGACTTTCAGAAGCTGCTAATGAAGCAAGATTTCAAAA	2996
Db	1345	AAATC-----TACTGAAGAGATTTTCTTAGTGATAGGTTGAGAA	1386
Qy	2997	CTTAAAGAGCTGTCCAAGCTCCCGGCATTTTATGATAGATTTTCAAGTTCGCTGGCTCCA	3056
Db	1387	CTAAAAGAGCTTTCGAAGTTCCTTGATATCTATGAAGATTTGACTAGATCATTTAGCTCCA	1446
Qy	3057	AGCATTTGGGAGCTTGAAGATTAATAAAGGTCCTTTTCCAGACTCTTTTGGTGGGAAG	3116
Db	1447	AACATATGGGAGTTGGATGATGTCAAAAGAGGTCTCTTTTCCAGCTTTTTCGGCGGTAAT	1506

Db 2584 CTGGCACCCTGATGACTGAAGCGCGGTGGTATCCATGGAGACAACGTGAGGAGATT 2643
QY 4257 TGA 4259
Db 2644 TGA 2646

RESULT 4
AAI67610
ID AAI67610 standard; cDNA; 723 BP.
AC AAI67610;
DT 27-FEB-2002 (first entry)
XX Cell cycle protein 2 (CC-2) partial cDNA sequence.
DE Cell Cycle Stress-Related Protein; CCSR; cell cycle protein; CC-1; CC-2;
KW CC-3; environmental stress; ss.
XX Physcomitrella patens.
OS WO200177354-A2.
PN 18-OCT-2001.
PD 06-APR-2001; 2001WO-US011294.
PF 07-APR-2000; 2000US-0196001P.
PR (BADI) BASF PLANT SCI GMBH.
PA Costa E SilvaO, Bohnert HJ, Van Thiel N, Chen R;
XX Sarria-Millan R;
DR WPI; 2002-049151/06.

Novel Cell Cycle Stress-Related Protein useful for increasing tolerance to environmental stress, is selected from Cell Cycle Proteins 1-3, or their orthologs.
Disclosure; Fig 1B; 90pp; English.
The invention relates to a Cell Cycle Stress-Related Protein (CCSRP), isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1 protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSR and encoding nucleic acids are useful for increasing tolerance to environmental stress selected from salinity, drought and temperature, in transgenic plants including monocot and dicot selected from maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato, tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao, tea, Salix species, oil palm, coconut, perennial grass and forage crops. The nucleic acid is also useful for identifying organisms e.g. P. patens in a mixed population of microorganisms. The nucleic acids are also useful for evolutionary and protein structural studies. The proteins and nucleic acids are useful as markers for specific regions of the genome. The present sequence represents a partial cDNA sequence of CC-2 from P. patens
SQ Sequence 723 BP; 204 A; 136 C; 201 G; 182 T; 0 U; 0 Other;
Query Match 15.5%; Score 673; DB 6; Length 723;
Best Local Similarity 97.1%; Pred. No. 1.1e-207;
Matches 696; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 2753 GAAGCTGAGATCGTATTGAGTACAGAGTTTTCAGGCCATGGCAGTTGCGATTG 2812
Db 3 GCACACGGAGATCGTATTGAGGTAAACAGGAGTTTTCAGGCCATGGCAGTTGCGATTG 62
QY 2813 TCCGAATCAACGAACATTACGACATTGTATAGACCTACATCGATTGGTGCACGTCAA 2872

Db 63 TCCGAATCAACGAACATTACGACATTGTATAAGACCTACATCCGATTCGTCACGTCAA 122
QY 2873 GAAGCTGACAGGGGTGACCTGCAAACTGAAGATCCCTATGAGATGGATAGGAGAAATGA 2932
Db 123 GAAGTCTGACAGGGGTGACCTGCAAACTGAAGATCCCTATGAGATGGATAGGAGAAATGA 182
QY 2933 TATGTATGCTGGGTATCATGAAAGTGATACCTTCAGAACTGCTAATGAAGCAAAAGATTCA 2992
Db 183 TATGTATGCTGGGTATCATGAAAGTGATACCTTCAGAACTGCTAATGAAGCAAAAGATTCA 242
QY 2993 AAAACTTAAAGAGCTGTCCAAAGCTCCCGGCAATTTATGATAGCTTTCAAGGTCCGCTGGC 3052
Db 243 AAAACTTAAAGAGCTGTCCAAAGCTCCCGGCAATTTATGATAGCTTTCAAGGTCCGCTGGC 302
QY 3053 TCCAAAGCAATTTGGGAGCTTGAAGATATTTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
Db 303 TCCAAAGCAATTTGGGAGCTTGAAGATATTTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362
QY 3113 GAAGCTTAAAGAAAATTTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGT 3172
Db 363 GAAGCTTAAAGAAAATTTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGT 422
QY 3173 TGGGACCTCGTACCAAGTAAATCTCAGTCTGCTTCAGTATGTGCAACAAGATAGCTCCTCG 3232
Db 423 TGGGACCTCGTACCAAGTAAATCTCAGTCTGCTTCAGTATGTGCAACAAGATAGCTCCTCG 482
QY 3233 TGGATCTACACTAGTGGCGGAGGAGTTCCGCGGTGGGCTGACAGCGTATGT-AAACA 3291
Db 483 TGGATCTACACTAGTGGCGGAGGAGTTCCGCGGTGGGCTGACAGCGTATGTAAACGA 542
QY 3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG 3351
Db 543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG 602
QY 3352 GGATATGCTGTATCGATGAGTTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db 603 GGATATGCTGTATCGATGAGTTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662
QY 3412 AGTAAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTTGCTGCTGGA 3468
Db 663 AGTAAATGGAGCAACAAACGGTATCTGGACCCAGCGGTTTCATGCTCGTGAAGCCGA 719

RESULT 5
AAD60810
ID AAD60810 standard; cDNA; 723 BP.
XX AAD60810;
DT 15-JAN-2004 (first entry)
XX Physcomitrella patens CC-2 EST (expressed sequence tag) cDNA.
DE Cell cycle stress-related protein; CCSR; metal; cell cycle protein; CC;
KW transgenic plant; environmental stress; stress tolerance; salinity;
KW drought; temperature; chemical; oxidative stress; expressed sequence tag;
KW EST; ss.
XX Physcomitrella patens.
OS US2003097675-A1.
PN 22-MAY-2003.
PD 06-APR-2001; 2001US-00828062.
PF 07-APR-2000; 2000US-0196001P.
XX (SILV/) COSTA E SILVA O.
PA (BOHN/) BOHNERT H J.
PA (THIE/) THIELEN N V.
PA (CHEN/) CHEN R.
PA (SARR/) SARRIA-MILLAN R.

XX Costa E SilvaO, Bohnert HJ, Thielen NV, Chen R, Sarria-Millan R;
XX WPI; 2003-765533/72.
XX Novel cell cycle stress-related protein capable of conferring stress
XX tolerance such as tolerance towards salinity, drought, temperature,
XX chemical, pathogens, to plants upon over-expression.
XX Example 5; Fig 1B; 62pp; English.
XX The invention relates to an isolated cell cycle stress-related protein
XX (CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
XX and its orthologues. Recombinant expression vector is useful for
XX producing a transgenic plant containing a CCSRP coding nucleic acid,
XX where expression of the nucleic acid in the plant results in increased
XX tolerance to environmental stress as compared to a wild type variety of
XX the plant which involves transforming a plant cell with the recombinant
XX expression vector, generating from the plant cell a transgenic plant with
XX an increased tolerance to environmental stress as compared to a wild type
XX variety of the plant. CCSRP is useful for conferring stress tolerance
XX such as tolerance towards salinity, drought, temperature, metal,
XX chemical, pathogens and oxidative stress or their combinations to plants.
XX The present sequence is Physcomitrella patens CC-2 EST (expressed
XX sequence tag) cDNA
XX
SQ Sequence 723 BP; 204 A; 136 C; 201 G; 182 T; 0 U; 0 Other;

Query Match 15.5%; Score 673; DB 10; Length 723;
Best Local Similarity 97.1%; Pred. No. 1.1e-207;
Matches 696; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 2753 GAAGCTCGAGATCGTATTGAGTAAACAGGAGTTTCAAGGCCATCGGAGTTTCGAGTTGG 2812
DB 3 GCACACGAGAGATCGTATTGAGTAAACAGGAGTTTCAAGGCCATCGGAGTTTCGAGTTGG 62
QY 2813 TCCGAATCAACGAACATTACGAGCATTTGTAAGACCTTACATCGATTCGTCGACGTCAA 2872
DB 63 TCCGAATCAACGAACATTACGAGCATTTGTAAGACCTTACATCGATTCGTCGACGTCAA 122
QY 2873 GAAGTCTGACAGGGTCGATCGAATCAACTGAATCTATGGAGTGGATAAGAGNATCA 2932
DB 123 GAAGTCTGACAGGGTCGATCGAATCAACTGAATCTATGGAGTGGATAAGAGNATCA 182
QY 2933 TATGTATGCTGGGTATCATGAAGTCAATCTTCAAGAGCTGCTTAATGAAGCAAGATTCA 2992
DB 183 TATGTATGCTGGGTATCATGAAGTCAATCTTCAAGAGCTGCTTAATGAAGCAAGATTCA 242
QY 2993 AAAAATTAAGAGCTGCTCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
DB 243 AAAAATTAAGAGCTGCTCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 302
QY 3053 TCAAGCATTTGGAGCTTGAAGATATTAAGAGGTCTTCTTGGCAGCTCTTGGTGG 3112
DB 303 TCAAGCATTTGGAGCTTGAAGATATTAAGAGGTCTTCTTGGCAGCTCTTGGTGG 362
QY 3113 GAAGGTAGAAATTCATCTGGAGCATCTTCCGAGTGACATCAATGTTTACTCTT 3172
DB 363 GAAGGTAGAAATTCATCTGGAGCATCTTCCGAGTGACATCAATGTTTACTCTT 422
QY 3173 TGGGGACCTGGTACCAAGTAAATCTCAGCTGCTTCAAGTATGTCACAGATGCTCCTCG 3232
DB 423 TGGGGACCTGGTACCAAGTAAATCTCAGCTGCTTCAAGTATGTCACAGATGCTCCTCG 482
QY 3233 TGGAAATCTACATAGTGGGCGAGAAAGTTGGCGGTGGGCTGACAGCGTATGT-AACGA 3291
DB 483 TGGAAATCTACATAGTGGGCGAGAAAGTTGGCGGTGGGCTGACAGCGTATGTAAACGA 542
QY 3292 AGGATCCAGAAATCCAGAGACGGTATTTGAGAGCGGAGCTTTGGTCTTAGTGATCGTG 3351
DB 543 AGGATCCAGAAATCCAGAGACGGTATTTGAGAGCGGAGCTTTGGTCTTAGTGATCGTG 602
QY 3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCATTAATGTCGCCGAGCATGCTTTCATG 3411

DB 603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCATTAATGCCGAGCATGCTTTCATG 662
QY 3412 AGGTAATGGAGCAACAAACCGGTATCTGTAGCCAAAGGGGGTATATTCCTCGCTGA 3468
DB 663 AGGTAATGGAGCAACAAACCGGTATCTGCAGCCCAAGCGGTTCATGCTCGTGAAGCGGA 719
RESULT 6
ADX36902
ID ADX36902 standard; cDNA; 1571 BP.
AC ADX36902;
XX 21-APR-2005 (first entry)
DT Plant full length insert polynucleotide seqid 19722.
DE
XX plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX Unidentified.
XX US2004034888-A1.
PN 19-FEB-2004.
PD 28-APR-2003; 2003US-00425114.
PF 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX New recombinant DNA construct, useful for improving plant tolerance to
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
XX pests, for conferring increased resistance to plant disease, or for
XX improving yield.
XX Claim 1; SEQ ID NO 19722; 15pp; English.
XX The invention describes a recombinant DNA construct comprising a
XX polynucleotide consisting of a sequence encoding an amino acid sequence
XX available in electronic form from the US patent office at
XX ftp.sequata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
XX of the invention are also useful in physical arrays of molecules and as
XX plant breeding markers. The recombinant DNA construct is useful for
XX improving plant tolerance to cold, heat, drought, herbicides, extreme
XX osmotic conditions, pathogens or pests, for manipulating growth rate in
XX plant cells by modification of the cell cycle pathway, for conferring
XX increased resistance to plant disease, for producing galactomannan,
XX lignin or plant growth regulators, for increasing the rate of homologous
XX recombination in plants, for improving yield by modification of
XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
XX or by providing improved plant growth and development under at least one
XX stress condition or for modifying seed oil or protein yield and/or
XX content. This sequence represents a plant full length insert
XX polynucleotide that can be used in the recombinant DNA construct of the

```
CC invention.
XX SQ Sequence 1571 BP; 437 A; 311 C; 403 G; 420 T; 0 U; 0 Other;
Query Match 14.6%; Score 634.2; DB 13; Length 1571;
Best Local Similarity 67.5%; Pred. No. 8.3e-195;
Matches 907; Conservative 0; Mismatches 433; Indels 3; Gaps 1;

QY 2917 TGGATAAGGAGAAATGATGATGCTGGGTATCATGAAGTGATGATCTTCAGAAAGCTGCTA 2976
Db 32 TGGAGGACACCATGGATATGATAATCTTAACGGTAGCAAAATCTACTGAAGAGGATTTTC 91
QY 2977 ATGAACAAAGATTCAAAACCTTAAGAGCTCTCCAAAGCTCCGGGGCATTTATGATAGAC 3036
Db 92 TTAGTGATAAGGTTGAGAAAGCTTAAAGAGCTTTTGAAGCTTTCGAAAGTTGCTGATATCTATGAAGAT 151
QY 3037 TTTCAAGGTCGCTGCTCCAAGCATTTGGGAGCTTGAAGATATTAAGAGGCTCTCTTTT 3096
Db 152 TGACTAGATCATAGCTCCAAACATATGGAGTTGGATGATGTCAAGAGAGGTCTCCTTT 211
QY 3097 GCCAGCTCTTTGGTGGAAAGGCTAAGAAAATTCATCTCGAGCATCTTTCCAGGTGACA 3156
Db 212 GCCAGCTTTTCGGGGCAATCCCTTGAAGCTTCTTGGAGCTAGTTTCCGGGGTGACA 271
QY 3157 TCAATGTTTTACTTTGTTGGGACCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGC 3216
Db 272 TCAATATTTTTACTTTGTTGGGGGACCTTGGAAACAAGTAAATCCCAGGCTTCTCCAGTACATGC 331
QY 3217 ACAAGATAGCTCTCGTGGAAATCTACACTAGTGGGGAGGAAGTTCCGGGGTTGGGCTGA 3276
Db 332 ATAAACTGTCTCTCGTGGTATCTATACAGTGGTAGAGGAAGTTCTGCTGTGTGCTTAA 391
QY 3277 CAGCGTATGTAAGAAAGGATCCAGAAACTCGAGAGACGGTATTTGGAGACGGAGCTTTGG 3336
Db 392 CTGCTTATGTTACCAAGACCCTGAGACTGGCGAAACTGTTCTAGAAAGTGGAGCACTTG 451
QY 3337 TTCCTTAGTATCGTGGGATATGCTGTATCGATGAGTTTCACAAAATGCTGTATATGCC 3396
Db 452 TTTTGAGTGACAAAGGTTGTTGTTGCTAGATGAGTTTGATAAGATGTCGTATATGCC 511
QY 3397 GAAGCATGCTTCATGAGTAAATCGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCA 3456
Db 512 GAAGCATGTTACAGAGGTGATGGAACAGAGCATGATCCATTCGCAAGGCTGGATAA 571
QY 3457 TTGCTCTCGTGAAACGCTCGAGCTGTCTCTTGATGTGCAAACTCTAGTGGGTCCGAT 3516
Db 572 TTGATCTTTAAACGCTAGACATCTGTCTGGCATGTGCAATCTCTACTGAATCACGTT 631
QY 3517 ACATCGCGCCCTTTCTGTGATTTGATAACATCCAGCTTCTTCCAACTCTACTTTAGAT 3576
Db 632 ACAATCCAAAGGCTCTCTGTAATTGACAAATCCACCTTAGCGCAACGCTACTTTCAAGAT 691
QY 3577 TTGATTTAATTTACTTAATGCTCGACAAACAGAGCAAAACGATCGTCTCTCGCCA 3636
Db 692 TCGACCTGATTTATCTTCTTTCGACAGGCGATGAGCAAACTGATAGCGCTCGCAA 751
QY 3637 GGCATCTCGTGGCTTTTACATATGAAAACTATGAAGTTTCAAAGCAGGACGCTTAGATC 3696
Db 752 AGCATATTGTTTGGTTCATTTTGAATCCAAATTTAGAGGAGCTCGAGTCTTGGACT 811
QY 3697 TACAAACACTTACCGGTATATCACCTATGCTCGTCAGCATGTATCTCTACATTTAAGTG 3756
Db 812 TGCAGACACTAGTTTCTCATAAAGCTATGCAAGGAAGTATATTCAGCCACAGATTTATCTG 871
QY 3757 ATGAAGCTCGAAGATTTGATTAATGGCTATCTTCAGATCGCCCAAAGGGCAACTTTC 3816
Db 872 ATGAAGCTCGAAGAGTTTAACTCGTGGCTATGTCGAGATGAGAAAAGAGGGAATAGCC 931
QY 3817 CTGGAAGCAGTAAAGGTTGATTAACAGCCACACTCGGCAACTCGAAAGTATGATTCGTA 3876
Db 932 CTGGGAGCAGAAAGAGTTCATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGTC 991
QY 3877 TCAGTGAGCCCTAGCTCGATCGATGATTTTCTGAAGTGGTAGAGAAAGTTGATGACGAG 3936

Db 992 TCAGGAAGCATTTAGCCCAATGCGGTCTCTGAAGTGGTCGAGGTGCGGATGTTGTGG 1051
QY 3937 AAGCTGTGCGCTTTTAGACGTGCGTTTGGCAGCAATCTGTACTGATCATGCAACAGGTA 3996
Db 1052 AGGCATTTCAGGCTTCTTGAAGTCGCAATGCGCAGCTCTGCGACGATCATGCAACTGGTA 1111
QY 3997 CGATAGACATGATCTTATCAGACTGAGTGTGCGCCAGCGAGGTATTGTCGCGGCCA 4056
Db 1112 CGATTGATGATGATCTGATCATGACGGGATATCCGCAAGCGAAAGGAGGAGCGGAGGA 1171
QY 4057 ACTTGTAGCTGCTCTGCGAGAGCTTATAGCAGATATAAATTTCACTTGGCAGCTCTCTG 4116
Db 1172 ACCTCGTTGCCGCAACCCGTAACCTGATTTGGAGAAAATGCAAGTTGGAGGCCCTC--- 1229
QY 4117 GCTTGAAGACAGTCAGCTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGAAGTGT 4176
Db 1230 -GATGCGCATGATTGAGTTGCTGGAGGAACTGAGGAAGCAGAGCTCAATGGAAATTCATA 1288
QY 4177 TGCAGGATATTAAAAATGCTCTGCGTAGCTTCCAAAGGAGAGGCTTTCTTACTGTCCATG 4236
Db 1289 TGCACGAACTCCGCGGTGCTCTTGGCACCTGATGACTGAAAGGCGCGTGTATCCATG 1348
QY 4237 GTGACATAGTCAAGAGAGTTTGA 4259
Db 1349 GAGCAACGTGAGGAGAGTTTGA 1371

RESULT 7
ADB69927
ID ADB69927 standard; DNA; 2970 BP.
XX AC ADB69927;
XX DT 04-DEC-2003 (first entry)
XX DE C. neoformans open reading frame SEQ ID NO:2332.
XX KW ds; gene; fungicide; gene therapy; infection.
XX OS Cryptococcus neoformans.
XX PN WO2003052076-A2.
XX PD 26-JUN-2003.
XX PF 17-DEC-2002; 2002WO-US040225.
XX PR 17-DEC-2001; 2001US-0341261P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Zamudio C, Eroshkin AM;
XX DR WPI; 2003-533017/50.
XX PT P-PSDB; ADB70288.
XX PS New nucleic acid, useful for preparing a composition for treating an infection caused by Cryptococcus neoformans.
XX CC Claim 2; SEQ ID NO 2332; 136pp; English.
XX CC The invention relates to a novel purified or isolated Cryptococcus neoformans nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by Cryptococcus neoformans. The present sequence represents a C. neoformans sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
```


[illegible]

PF 26-SEP-2002; 2002US-00260238.
XX
PI (BUDWORTH P. MOUGHAMER T. BRIGGS SP. COOPER B. GLAZEBROOK J.;
PI MOUG/) MUGHAMER T. BRIGGS SP. COOPER B. GLAZEBROOK J.;
PI (BRIG/) BRIGGS S P. COOPER B. GLAZEBROOK J.;
PI (COOP/) COOPER B. GLAZEBROOK J. KREPS J. PROVART N. RICHKE D. ZHU T.
PA (GLAZ/) GLAZEBROOK J. KREPS J. PROVART N. RICHKE D. ZHU T.
PA (GOFF/) GOFF S A. KREPS J. PROVART N. RICHKE D. ZHU T.
PA (KATA/) KATAGIRI F. KREPS J. PROVART N. RICHKE D. ZHU T.
PA (KREP/) KREPS J. PROVART N. RICHKE D. ZHU T.
PA (PROV/) PROVART N. RICHKE D. ZHU T.
PA (RICH/) RICHKE D. ZHU T.
PA (ZHU/) ZHU T.
XX
PI Budworth P. Moughamer T. Briggs SP. Cooper B. Glazebrook J.;
PI Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2004-190374/18.
XX
XX New rice promoter, useful for manipulating crop plants to alter or
PT improve phenotypic characteristics, e.g. produce large quantities of oil
PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT or high nutritional value.
XX
XX Claim 70; SEQ ID NO 580; 230pp; English.
XX
XX The invention relates to plant nucleotide sequences that direct seed-,
CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC or constitutive transcription of an operatively linked nucleic acid
CC segment. The invention also relates to a method for augmenting a plant
CC genome and a method of identifying a gene, where its expression is
CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC encode are useful for manipulating crop plants to alter or improve
CC phenotypic characteristics, to produce large quantities of oil or
CC proteins, to incur resistance to insecticides, viruses or fungi, and to
CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC have a high nutritional value with reduced apical dominance or dwarfism,
CC early flowering or altered metabolic pathways. This sequence represents a
CC plant nucleic acid of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification but was obtained in
CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1723 BP; 459 A; 373 C; 461 G; 429 T; 0 U; 1 Other;
SQ

Query Match 11.3%; Score 490; DB 12; Length 1723;
Best Local Similarity 69.4%; Pred. No. 9.4e-148;
Matches 762; Conservative 0; Mismatches 285; Indels 51; Gaps 5;
2989 TTCAAAACCTTAAGAGCTGTCCAAGCTCCGGGCAATTTATGATAGACTTTCAAGTCCG 3048
606 TTCAATNAGAGGAGGTGTCAAGTTGCCAGATATATGACAGATTAACATAGGTCTAT 665
3049 TGGCTCCAAGCAATTTGGGAGCTTGAAGATATTAAGAGGCTCTCTTTTCG 3105
666 TGGCTCCAACATATGAGGAGCTGGACGACGTTTAAAGGGGCGCTCTTTCCAGCAGCTTT 725
3106 TTGGTGGGAGGCTAAGAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTT 3165
726 TTGGTGGAAATGCTTTGAGGCTTCTCTGGAGCTAGTTTCCGAGGCGACATCAATATT 785
3166 TACTTGTGGGAGCCCTGGTACCAGTAATCTCAGCTGCTTCAGTATGTCACAGATAG 3225
786 TGTCTTGTGGTATCTGGAAACAGTAATCCAGCTTCTCCAAATACATGCAACAACTGT 845
3226 CTCCTCGTGAATCTACACTAGTAGGGGAGGAGTTCCGGCGGTGGGCTGACAGCGCTATG 3285

Db 846 CTCCTCGTGGCATTATTATACAAAGTGGCAGAGGAAGTTCAGCTGTGTGGCCTTACTGCATACG 905
QY 3286 TAACGAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTCTTAGTG 3345
Db 906 TTACCAAGGATCCCTGAACCTGGTGAACCTGTTTGAAGAGTGAGCGCTTGTGTTGAGTG 965
QY 3346 ATCGTGGGATATGCTGTATCGATGAGTTGCGACAAAATGCTCTGATATATGCCCAAGCATGC 3405
Db 966 ACAAGGTTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1025
QY 3406 TTTATGAGGTAAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTTGCTCGC 3465
Db 1026 TACATGAGGTGATGGAACAAACAGACTGCTCCATTTGCCAAGGCTGGAATAATTTGCACTTT 1085
QY 3466 TGAAGCTCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3525
Db 1086 TAAATGCCAGAACATCAGTTCTTAGCATGTGCAAAATCCTACTGAATCAGCTATTAATCCAA 1145
QY 3526 GCCTTTCTGTGATGATAAATCAGCTTTCTCCAACTCTACTTTCTAGA--TTTGATTT 3583
Db 1146 GGTCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1205
QY 3584 AATTACTTAATGCTCGACAAACCCAGACGAGCAAAACGATCGTCTGCGCAGGATCT 3643
Db 1206 CATTTATCTGATATTGGACAAGGAGATGAGCAAACTGATAGACGCTGGCTAAGCATAT 1265
QY 3644 CGTGCTTTACATGATAAATCTATGAAGTTTCAAGCAGGAGCGCTTAGATCTCAAAAC 3703
Db 1266 TGTTCGTGCTATTTTGAGAA----- 1286
QY 3704 ACTTACCGGTATATACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3763
Db 1287 ---TCCAACTTACATPAAGTTATGCAAGGAAGCATATACACCAAGTTATCTGATGAAGC 1343
QY 3764 TGCTGAAGATTTGATTAATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3823
Db 1344 TGCAGAAGATTGACTCGCGCTATGTTGAGATGAGGAAGAAAGAGGAAACAGCCCTGGTAG 1403
QY 3824 CAGTAAAAGGTGATTAACAGCCACACCTCGGCAACTCGAAAGATGATGATGATGATGATGATGAT 3883
Db 1404 CAG---AAAGAGATTAACCTGCGACAGCTCGACAAATGAGAGCTTGAATTCGGCTCAGTGA 1460
QY 3884 AGCCCTAGCTCGAATGAGATTTCTGAAAGTGTGAGAAAGTTGATGACAGCAAGAGCTCT 3943
Db 1461 AGCACTGCGCCCAATGCGGATTTCTGAAATGTTGAAGTACAAAGATGTTGTAGAGGCTTT 1520
QY 3944 GCGCCTTTTAGACGCTGCTTTGCGCAATCTGCTACTGATCATGCAAC-AGGTACGATAG 4002
Db 1521 CAGGCTTCTCGAAGTTGCCATGCGCAATCGCACTGATCATGCTGCGGTACATCG 1580
QY 4003 ACATGGATCTTATCAGACTGGAGTGTGCGCAGGAGGCTATTCGTGCGGCCAATTCG 4062
Db 1581 ATATGGATCTTATCATGACTGGAATATCTGCGAGCGAAAGGAGGAGCGGAGCCAGCTTGC 1640
QY 4063 TAGCTGCTCTCGAGAGC 4080
Db 1641 TGGAAAGAAATTAGGAAGC 1658
RESULT 9
ABZ32193
ID ABZ32193 standard; DNA; 2733 BP.
XX
AC ABZ32193;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential gene SEQ ID NO 6480.
XX
KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX

Db	2191	TATGCTAAAGAGAAATTTCAATCCCGTGATGACAGAAGAAGTAAAAATGAATAGTTTAGA	2250
Qy	3783	GGCTACTTGTGAGATGCGCCAAAGGGCAACTTTCTCGAAGCAGTAAAAAGGTGATAACA	3842
Db	2251	TCATACGTTGAAATTGAGAAAATTTGGGGAGGATGCTAGATCATCAGAAAAAGAGAATCACC	2310
Qy	3843	GCCACACTCGGCAACTCGAAAGTATGATTGTTATCAGTGAAGCCCTAGCTCGAATCGAGA	3902
Db	2311	GCTACTACAAGACAAATTTGGAAATCGATGATTAGATTGTCTGAAGCTCATGCTAAATGAGA	2370
Qy	3903	TTTTTCTGAAGTGGTAGAGAAAAGTTTGATCGACGACAAGCTGTGCGCCCTTTTAGACGTGCGT	3962
Db	2371	TTATCTGAAAGAGTTGTAATTCATCGATGTTAAAGAAGCAGTAGAATTAATTAATCGGCC	2430
Qy	3963	TTGCAGCAATCTGCTACTGATCATGCACAAGGTACGATAGACATGGATCTTATCACGACT	4022
Db	2431	ATTTAAGATTATGCTACTGATCCAGTTACTGGTAGAATCGATATGGATATGATTCAAACT	2490
Qy	4023	GGAGTGTGCGGCCAGGACGGTATTTCGTGCGGCCAACTTTGCTAGTGCTCTTCGCGAGAGCTT	4082
Db	2491	GGTACTACGGCACACAAGAAGAGTACAGAAGATTTGGTCACTGAAATTTATGAAAATC	2550
Qy	4083	ATACGATATAAAATTT	4097
Db	2551	ATCGAAGAAAATAAT	2565

RESULT 10
ADT47760
ID ADT47760 standard; cDNA; 2802 BP.

02-DEC-2004 (first entry)

DE Bacterial polynucleotide #22511.

Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene: ss.

OS Bacteria.

AA
PN
US2003233675-A1.XX
PD
18-DEC-2003

20-FEB-2003: 2003US-00369493.

XX
PR 21-FEB-2002: 2002US-0360039P.

PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.
PA (STAT/) STATER S C.

PA (CHEN//) CHEN X.
PA (GOLD//) GOLDMAN B S

XX
DT
Cao v Hinkle

XX
DB
WPT: 2004-061375/06

XX
PT New recombinant DNA construct comprising a promoter

PT for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

PS Claim 1: SEQ ID NO 46198; 122pp; English.

CC The invention relates to a recombinant DNA construct comprising a

CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

Qy	2382	CTTTTCAATCTCAAAGCATCGGTGCACATCGTGAACTCAACCTTCAGATATAGACAAA	2444
Db	724	CTTTTGGTCTCGACTCGACTGTGAAATAGAGAGACCTCGATCTCTGCAGACATGGATAAA	783
Qy	2442	TTGGTTCTCTTTAAAGGAATGGTTATCCGGTGCAGTTCATATCATACCTGAATTAAGGGG	2501
Db	784	CTAGTAAGACATTAAGGGCTTAGTCAATTCGAAACGACACCACTCATCTCTGATATGAAGAA	843
Qy	2502	GCCTTCTTCAAAATGTTTATGTGTGGTGCACTCGCTCCGCTAGTTTACAGTTGTTAAAGGG	2561
Db	844	GCCTTCTTCGTTGCCAAGTCTGCAACCATGTGGTTCAGTTCAGATCGACA---TTGATCGTGA	900
Qy	2562	CGGGTTGAGGACCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGAAATGCTATGTCT	2621
Db	901	AAGATTGCGGAAACCACTAGTGTGCCACGTCCAGTGTGTAAGGAAACGAACTCGATGCAA	960
Qy	2622	CTTATTCAAAATCGATGCATTTTGTGCAAAATTAAGCAGATAGTGGCTCTTCAAGAAAATCTCCA	2681
Db	961	CTCATCTCAATCAACGGCTGTGTATTTTGGCCGACACAGCAGGTCTCAAGTTGTCAGGAAACACCT	1020
Qy	2682	GATGCCAATCTCTGAAGCAGAGACTCCACACACAGTTCAGCATGTGTTTATACAACTATG	2741
Db	1021	GACAGCAATCTGATGCGCAGACTCTCTCACTCGGTTTCCCTTTGTGTGTATGATGAGCTG	1080
Qy	2742	GTTGATGCTGTGAAGCCTGGAGATCGTATTTGAGGTAAACAGAGATTTTCAAGGCCATGGCA	2801
Db	1081	GTGGATGCTGCAAGGCTGGTGATCGGGTCCGAGTGCAGCGTATTTTCCGGTGCACCCCT	1140
Qy	2802	GTTTCGATTGGTCCGAATCAACGAAATTAAGAGCATTTGATTAAGACCTACATCGATTCG	2861
Db	1141	GTGCGCTTAACTCTCGCCACGCTACACAGAAAGTCGCTGTTTCAAGACGTACATAGATGTT	1200
Qy	2862	GTGCACGCTCAAGAACTGTGACAGGGTCGACTGCAAACTCAAGATCCTATGGAGATGGAT	2921
Db	1201	CTTCATGTTTCAGAAAGATCGATCGCAAGAGTTGGGTATCGACGCTCTCGAACATCGAGCAG	1260
Qy	2922	AAGGAGAATGATATGTATGTCTGGGTATCATGAAAGTGTATACTTCAGAAAGCTGCTAATGAA	2981
Db	1261	GAGCTCTCGNAACAGCGGCTGGGGATGCAGNACAAACACGTAGGCTCATCTGCGAGGAG	1320
Qy	2982	GCAAGATTCAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTTATGATAGACTTTCA	3041
Db	1321	GAAGAG-----AAAAATTAAGCGAACTGCTACCAAGACTGATCTGTATGAGCTTCTCTCT	1374
Qy	3042	AGGTGCGTGGCTCCAAGCAATTTGGGAGCTTGAAGATATTAAAGGCTCTCTTTGCCAG	3101
Db	1375	CGGTCTTTGGCCCCCAGCATCTACGAGATGGAACGACGTGAAGAGGGAATCCTGCTTCAG	1434
Qy	3102	CTCTTTGGTGGAGGGCTAAGAAAATTTCCATCTCGAG-----CATCTTTCCGAGGTGCAC	3155
Db	1435	TTGTTTGGAGGCCAACACAGACCTTCCAGNAGGTTGGTAACCCACGATACCGTGGAGAT	1494
Qy	3156	ATCAATGTTTACTTGTTTGGGACCCCTGGTACCAAGTAAATCTCAGCTGCTTCAGTATGCT	3215
Db	1495	ATCAATATCTTCTCTGTGTGTGACCCCATCTACATCCAAGTCCCAGCTTCTTCGTTCAGTCT	1554
Qy	3216	CACAAGATAGCTCTCTGTTGGAATCTACACTAGTGGGCGAGGAAGTTTCGGCGGTTGGGCTG	3275
Db	1555	CATAAGATTTGCCCTTCGCGGTGTGTATACCAAGCGCAAGGGCTCCTCGGCTGTTGGTCTT	1614
Qy	3276	ACAGCTATGTAAACGAAGGATCCAGAAAATCTCGAGAGCGGTATTGGAGAGCGGAGCTTTCG	3335
Db	1615	ACGGGTACGTACACCGCGATCTCTGAAAACCGCCAGATGTCTCTCGAGTCGGGTGCGCTTG	1674
Qy	3336	GTTCTTATGATGATCGTGGGATATGCTGTATCGATGAGTTTCGACAAAATGTCTGATAATGCC	3395
Db	1675	GTTCTTTACAGACGGCGTATCTGTTGTGATCGACGAGTTTCGACAGATGAACGAATCCACT	1734
Qy	3396	CGAAGCATGCTTCATGAGGTAAATCGAGCMAAAACGGTATCTGTAGCCAAAGGGGGTATC	3455
Db	1735	CGGTCCGTTCTGCAATGAAGTCAATGGAACACAGACAGTATCTATTCGCAAGGCGGCAATT	1794
Qy	3456	ATTGCCCTCGGTGAACGCTCGACAGCTGTGTCTCTTCGATGTGCAAAATCTAGTGGGTCCCGA	3515

[illegible]

RESULT 12
ABT21208
ID ABT21208 standard: DNA; 3072 BP

AA	Aspergillus fumigatus essential gene #3566.
DE	
XX	
KW	Fungicide; cytostatic; essential gene: Aspergillus fumigatus; infection;
KW	cancer; contamination; biofilm; antibody; immune response; ds.
XX	
OS	Aspergillus fumigatus.

(ELIT-) ELITRA PHARM INC.

Db 1172 --CTGCCACACCCACAGCATGGCACTCATCCAAACCGCTCCCTCTCTCTGACAAG 1229
Qy CAGATAGTGGTCTTCAAGAACTCCAGATGCCATTCCTGAGGAGAGACTCCACACACA 2714
Db 1230 CAGATGATCAAGCTTTCAGGAGTCTCCGGAAGATCGCTGCGAGGGCAGACACACACA 1289
Qy 2715 GTGAGCATGTGTTTATACACACATATGTTGTGCTGTGAAGCCTCGAGATCGTATTGAG 2774
Db 1290 GTTATCTGTGCTCACAATGATCTCGTTGACAAGTCCAGCTCGGGACAGAGTGAAT 1349
Qy 2775 GTAAAGAGGATTTTCAAGGCCATGCGCAGTTTCGAGTTGGTCCGAATCAACGAAATTAACA 2834
Db 1350 GTTACAGGCATCTATCGAGCTGTCCTATTCGAGTCAATCCAAAGATGAGTATGTAAG 1409
Qy 2835 GCATTGTATAAGACCTACATCGATTGGTGCACTCAAGAGTCTGACAGGGGTCCAGTG 2894
Db 1410 TCTGTCTACAAAACCCACATTTGATGTCATTATTCGGAACCGGATGCAAAACCTCTG 1469
Qy 2895 CAAACTGMAATCCTATGGAGATGGATAAGGAGATGATATGATGCTGGGTATCATGAA 2954
Db 1470 CA-----TGCCCTTGATGAA 1484
Qy 2955 AGTGATACTTTCAGAGCTGCTAATGAAGCAAGATTCAAAACTTAAAGAGCTGTCCAAG 3014
Db 1485 GAAGCAGACAGAAAATTTTTCAGAGAACGTGTGGAATGCTTAAAGGAATTTTCCAGG 1544
Qy 3015 CTCCTGGGCAATTTATGATAGACTTTTCAAGGTCTGCTGCTCCAAAGCATTTTGGGAGCTTGAA 3074
Db 1545 AAACACAGACATTTATGAGAGCTTGTCTCAGCCCTGCTCCAAAGCATTTATGAACATGAA 1604
Qy 3075 GATATTAAAGAGGTCTTTTCCAGCTCTTTGGTGGGAGGCTTAAGAAATTT----- 3128
Db 1605 GATATAAAGAGGGAATTTTGTCTCAGCTCTTTGGGCGGCAAGGAAGGATTTTAGTCAC 1664
Qy 3129 CCATCTGGAGCATCTTTCGAGGTGACATCAATGTTTACTGTGTTGGGACCTCGTATACC 3188
Db 1665 ACTGGAGGGGCAATTTTCGGGTGAGATCAACATCTTGTGTGGGACCTCGTATACC 1724
Qy 3189 AGTAAATCTCAGCTGCTCAGTATGTCACAAAGATAGCTCTCTCGTGAATCTACATAGT 3248
Db 1725 AGCAAGTCCAGCTGCTGAGTACAGTGTACAACTCGTCCCGAGGGGCCAGTACACGTCT 1784
Qy 3249 GGGGAGGAGATTTGCGGGGTGGCTGACAGCGTATGTAACGAAGGATCCAGAACTCGA 3308
Db 1785 GGGGAAGGGCTCCAGTGCAGTTGGCTCACTGCGTACGTAAATGAAGACCCCTGAGACAAG 1844
Qy 3309 GAGACGCTATTGAGAGCGGAGCTTTGGTCTTGTAGTATCGTGGATATCTGTATCGAT 3368
Db 1845 CAGCTGGTCTGAGACAGAGTGTCTTGTCTGAGTGACAAAGGATCTGCTGTATCGAT 1904
Qy 3369 GAGTTGACAAAAATGCTGTATAATGCCGGAAGCATGCTTCATGAGGTAAATGGAGCAACAA 3428
Db 1905 GAGTTGCAAGATGAATGAAGTACAAGATCGGTATGATGATGATGATGATGATGATGATGAT 1964
Qy 3429 ACAGTATCTGAGCAAGGGGGTATCATTCCTCGCTGCAAGCTCGGACGTCTGTCTTT 3488
Db 1965 ACTCTGTCCATTTGCAAGGGTGGGATCATCTGTGAGTCAATGCGGCACTCTGTCTGT 2024
Qy 3489 GCATGTGCAATCTTGTGGTCCGATACAATGGCGGCGCTTTCTGTGATGATGATGATGATGAT 3548
Db 2025 GCAGCAGCAATCCCATTTGATGCTCTCAGTGGAAATCTTAAAAAACAACCATTTGAAACATC 2084
Qy 3549 CAGCTTCTCCACTTACTTCTAGATTTGATTTTAAATTTTAAATGCTGCAACAAACCA 3608
Db 2085 CAGCTGCTCATATTTTATCAAGGTTTGAATTTGATTTCTTCTCATGCTGAGCCCTCAG 2144
Qy 3609 GACGAGCAAAACGATCTGCTCGCAGGCGATCTCGTGGCTTTTACACTATGAAACATAT 3668
Db 2145 GACGAGCCCTATGACAGCGCTCTGGCTCACCACTGCTGCGACTACTACAGAGCGAG 2204
Qy 3669 GAAGTTTCAAAGCAGGAGCGCTTAGATCTTACAAACACTTACCGGCTATATCACTATGCT 3728

Db 2205 GAGCAGGCGAGAGAGAGCTCTTGACATGTCGCGGTGCTTAAAGCACTACATTCGCTTACGGG 2264
Qy 3729 CGTCAGCATGTACATCTTACATTAAGTGTATGAAGCTGCTGGAAGATTTGATTAATGCTAT 3788
Db 2265 CACAGCACCATCATGCGCGGCTAGTGAAGAGCCAGCAGGCTCTCATCGAGGCTTAT 2324
Qy 3789 GTTGAGATGCGCAAAAGGGCAACTTTCTCTGGAAGCAGTAAAGAGGTGATTAACAGGCACA 3848
Db 2325 GTAGACATGAGGAAGATTGGCAGTAGTACCGGGAA-----TGGTTCCTGCATAC 2372
Qy 3849 CCTCGCACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCT 3908
Db 2373 CCTCGACACTAGAGTCATTAATCCGCTTAGCAGAAAGCCCATGCTAAAGTAAAGTTGTCT 2432
Qy 3909 GAAGTGTGAGAGAAAGTTCATGTCAGCAGCAGAGCTGTGCGGCTTTTAGACGCTCGCTTTGCAG 3968
Db 2433 AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAAGGCTCCATCGGGAAGCTCTGAAG 2492
Qy 3969 CAATCTGCTACTGATCANGCAACAGGTACGATAGACATGATCGATCTTATCAGACTGAGTG 4028
Db 2493 CAGTCTGCAACTGATCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGGGATG 2552
Qy 4029 TCGGCCAGCAGAGGTATTCGTGCGGCCAACTTGTAGCTGCTCTGCGAGAGCTTATAGCA 4088
Db 2553 AGTGCCACCTCTCGTAAACGGAAGAAATTAGCTGAAGCAATTGAAAAAGCTTATTTTA 2612
Qy 4089 GATAAATTTTCACTTGGCAGCTCTCTGCTTGAAGCAGTCTGCTTCTTGAAGGATATC 4148
Db 2613 TCTAA-----GGGCAAAACACAGCTCTTAAATATCACGCACTTTTGAAGATAT 2663
Qy 4149 CGGAGCAAAAGCAGTGTGACGCTTGTGTCAGGATATTTAAAGATCTCTGCGTAGCCTC 4208
Db 2664 CGGGACATCTGACATAGCAATTAAGATATCTTGAAGAGCACTGCGTGCCTG 2723
Qy 4209 CAAGGAGAGGCTTTCTTACTGTCCATGTTGACATAGT 4246
Db 2724 GCAGATGATGATTTCTGACAGTACTGGGAAGACCGT 2761

RESULT 14

ABL65666

ID ABL65666 standard; DNA; 3273 BP.

XX ABL65666;

XX 15-MAY-2002 (first entry)

XX Lung cancer related gene sequence SEQ ID NO:4003.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.

XX Homo sapiens.

XX WO200194629-A2.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US010838.

XX 05-JUN-2000; 2000US-0209473P.

XX 05-JUN-2000; 2000US-0209531P.

XX 18-SEP-2000; 2000US-0233133P.

XX 18-SEP-2000; 2000US-0233617P.

XX 20-SEP-2000; 2000US-0234009P.

XX 20-SEP-2000; 2000US-0234034P.

XX 20-SEP-2000; 2000US-0234052P.

XX 22-SEP-2000; 2000US-0234509P.

XX 22-SEP-2000; 2000US-0234567P.

XX 23-SEP-2000; 2000US-0234923P.

XX 25-SEP-2000; 2000US-0234924P.

1845 CAGCTGGTCTGCAGACAGGCTCTTCTCTGAGTGAACGGCATCTGCTGATCAT 1904
3369 GAGTTTCGACAAATGTCGATAATGCCGGAAGCATGCTTCATGAGGTAAATCGACAA 3428
1905 GAGTTTCGACAAATGTAATGAAGTACAAGATCGGTATTGCAATGAATCATGGAACAGCAG 1964
3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTCCTCGCTGACGCTCGGAGCTCTGCTCTT 3488
1965 ACTCTGTCCATTGCAAGGCTGGGATCATCTGTGAGTCAATGCGCGCACCCTCTGCTCTG 2024
3489 GCATGTGCAATCCCTAGTGGTCCCGATACATCAATGCGGCGCTTCTGTGATTGATAAATC 3548
2025 GCAGCAGCAATCCATTCAGTCTCAGTGGATCTTAARAAAACAACCAATTGAACATC 2084
3549 CAGTTCTCTCAATCTTCTTCTAGATTTGATTAAATTAATTAATGCTCGACAAACCA 3608
2085 CAGCTGGCTCATCTTTATATCAAGGTTTGAATTTGATCTTCTCATGCTGACCCCTCAG 2144
3609 GACGAGCAAAACGATCTGCTCTCGCAGGATCTCTGGGCTTTACATATGAACATAT 3668
2145 GACCAAGCTATGACGCGCTCTGGCTCACCATCTGCTGCACTGTACTACGAGCGAG 2204
3669 GAAGTTTCAAGCAGGAGCGCTTAGATCTACAAACACTTACCGGCTATATCACCTATGCT 3728
2205 GACGAGCAGAGGAGGCTCTCTGACATGCGGCTTAAGGACTATCTGCTTACGCG 2264
3729 CGTCAGCATGTACATCTTACATTAAGTGAAGCTCTGGAAGTTTGAATTAATGGCTAT 3788
2265 CACAGCAACCATCATGCGCGCTAAGTGAGGAGCCAGCGCTCTCATCGAGGCTTAT 2324
3789 GTTGAGATGCGCAAAAGGCACTTCTCTGGAAGCAGTAAGAGGTGATAACAGCACCA 3848
2325 GTAGACATGAGGAAGTTGGAGTAGTGGCGGAA-----TGGTTTCTGCATAC 2372
3849 CCTCGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTCT 3908
2373 CTTGACAGCTAGAGTCATTAATCCGCTTAGCAGAGCCCATGCTAAAGTAAGATTTGCT 2432
3909 GAAGTGTAGAGAAAGTTGATGACAGCAGAAAGCTGTGCGCCTTTTGAAGCTGCTTTCAG 3968
2433 AACAAAGTTGAAGCATTGATGTGGAAGAGCCCAAGCCCTCCATCGGAAGCTCTGAAG 2492
3969 CAATCTGCTACTGATCATGCAACAGGTAGCATGATGATCTTATCAGACTGGAGTG 4028
2493 CAGCTGCAACTGATCCCGGACTGGCATCGTGACATATCTATTCTTACTACGGGATG 2552
4029 TCGGCCAGCGAGGCTATTCTCGTGGGCCAACTTGTAGTCTCTCGGAGAGCTTATAGCA 4088
2553 AGTGCCACCTCTCTGAACCGAAAGAAAGATTAGCTGAAGCATTTGAAAAGCTTTATTTA 2612
4089 GATAAAATTTTCACTGGCAGCTCTCTGGCTTGAAGACCAAGTCAAGCTTCTTGAGGATATC 4148
2613 TCTAA-----GGCAAAACACCACTCTAAATACAGCACTTTTGAAGATATT 2663
4149 CGGAGCAAAAGCAGTGTGGAGCTTGTGAGTTTGCAGGATATTAATAATGCTCTGGGTAGCCTC 4208
2664 CGGGGACAACTGACATAGCAATTAATAAGATATGTTGAAGAGCACTGCGGTGCGCTG 2723
4209 CAAGGAGAGGCTTCTTACTGCTCCATGTTGACATAGT 4246
2724 GCAGATGATGATTCTTCTGACAGTGAAGGAGACCGT 2761

RESULT 15
ABL65259
ID ABL65259 standard; DNA; 3273 BP.
XX
AC ABL65259;
XX
DT 15-MAY-2002 (first entry)
XX
DS Lung cancer related gene sequence SEQ ID NO:3596.

XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
cystostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
gene; ds.
OS Homo sapiens.
XX WO200194629-A2.
XX 13-DEC-2001.
PD 30-MAY-2001; 2001WO-US010838.
PF 05-JUN-2000; 2000US-0209473P.
XX 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-023133P.
PR 18-SEP-2000; 2000US-0231617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 26-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 28-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
PA (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
agent to be tested for anti-neoplastic activity, and determining a change
in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 3596; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
neoplastic agent. The method involves exposing cells to a chemical agent
to be tested for anti-neoplastic activity, determining a change in

Db	2553	AGTGCCACCTCTCGTAAACGAAAGAGAAATTAGCTGAGCAATTGAAAAAGCTTATTTA	2612
Qy	4089	GATAAAATTTACCTGGCAGCTCTCTGGCTTGAAGACCAAGTCAGCTTCTTGAGGATATC	4148
Db	2613	TCTAA-----GGGCAAAACACCAAGCTCTAAATATACCAGCACTTTTGAAGATATT	2663
Qy	4149	CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGTAGCCTC	4208
Db	2664	CGGGGACAACTGACATAGCAATTACTTAAAGATATGTTGAAGAAGCACTGCGTGCCCTG	2723
Qy	4209	CAAGGAGAGGCTTTCTTACTGTCCATGGTGACATAGT	4246
Db	2724	GCAGATGATGATTTCTCTGACAGTACTGGGAAGACCGT	2761

Search completed: December 6, 2005, 02:17:03
 Job time : 1616 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2005, 01:47:40 ; Search time 504 Seconds
(without alignments)
15334.996 Million cell updates/sec

Title: US-10-768-511-5
Perfect score: 4348
Sequence: 1 atggcgccgcactcaagt.....gcgaattccagagctgcg 4348

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues
Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4348	100.0	4348	3	US-09-828-062-5
2	673	15.5	723	3	US-09-828-062-2
3	209.6	4.8	2533	3	US-09-604-605-1
4	209.6	4.8	2533	3	US-10-101-080-1
5	198.6	4.6	1635	3	US-09-248-796A-4534
6	197.6	4.5	1488	3	US-09-248-796A-5904
7	186.4	4.3	2917	3	US-09-949-016-2468
8	183.8	4.2	2533	3	US-09-949-016-5021
9	180	4.1	3000	3	US-09-949-016-2621
10	173.2	4.0	996	3	US-09-248-796A-4570
11	160.8	3.7	3378	3	US-09-964-899-48
12	160.8	3.7	3445	3	US-09-976-594-323
13	157.6	3.6	3379	3	US-09-220-132-12
14	137	3.2	294	3	US-09-313-294A-4610
15	135.6	3.1	311	3	US-09-313-294A-83
16	133.8	3.1	297	3	US-09-313-294A-4978
17	127.8	2.9	12932	3	US-09-949-016-14363
18	123.2	2.8	507	3	US-09-401-064-158
19	120.2	2.8	519	3	US-09-248-796A-5916
20	105.2	2.4	1305	3	US-09-248-796A-4711
21	96.4	2.2	256	3	US-09-313-294A-2519
22	78.4	1.8	492	3	US-09-248-796A-5915
23	73	1.7	273	3	US-09-313-294A-1684
24	70.6	1.6	1664976	3	US-08-916-421B-1

c 24

ALIGNMENTS

RESULT 1

US-09-828-062-5
; Sequence 5, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNETT HANS J.
; APPLICANT: VAN THIESEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/09/828,062
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4348
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-062-5

Query Match 100.0%; Score 4348; DB 3; Length 4348;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCGCGCAGTCACTGAGGAAATGACCTCTTGTCTGCGACGGTTCATCTT 60
Db 1 ATGCGCGCGCAGTCACTGAGGAAATGACCTCTTGTCTGCGACGGTTCATCTT 60
Qy 61 TTTGGTTTTAGTTTGCATCTTGTGAGTTGAGAAAGGCGGTTCGTGCTCT 120
Db 61 TTTGGTTTTAGTTTGCATCTTGTGAGTTGAGAAAGGCGGTTCGTGCTCT 120
Qy 121 TGAGGTGTTCTTCTGATTTCTGATGAAATATGATGACTTGACATTTGAGCCG 180
Db 121 TGAGGTGTTCTTCTGATTTCTGATGAAATATGATGACTTGACATTTGAGCCG 180
Qy 181 TGTGTCGCCCATATCTTTCGCAATCTGAGGAGTGTCTACGCCATTCCCGCAAGTAACAT 240
Db 181 TGTGTCGCCCATATCTTTCGCAATCTGAGGAGTGTCTACGCCATTCCCGCAAGTAACAT 240
Qy 241 CACCGAGTTCGACATGACGCTCACCCGTGCGGCGGAGGCGGTACGGCAGACCC 300

Db 241 CACCGAGCTTCGACAAATCGAGCCTCAACCGTGGCCGGGGGAGGGCCGTACGCGAGACCC 300
Qy 301 CTACATCTCAGTTCCGAAGGAGAGGAGAGAAACGGATTCCGCTCGTCTAGGAGAGTC 360
Db 301 CTACATCTCAGTTCCGAAGGAGAGGAGAGAAACGGATTCCGCTCGTCTAGGAGAGTC 360
Qy 361 GATCTCGCAGTTTAGGCAATCTGTATTATAGTTTCCCTTTACGATGCGGGGACTCTCTGGAA 420
Db 361 GATCTCGCAGTTTAGGCAATCTGTATTATAGTTTCCCTTTACGATGCGGGGACTCTCTGGAA 420
Qy 421 CTCCTGGAACTCAGTGGCTATCTCGGTTTACGTTACCCAGTCGGGTACACCTATGGTA 480
Db 421 CTCCTGGAACTCAGTGGCTATCTCGGTTTACGTTACCCAGTCGGGTACACCTATGGTA 480
Qy 481 CCCCATCGTTCCATCGTGGCAGCCACAGTACAAACAGCGCAGTGAGCTTGGTCCCAGG 540
Db 481 CCCCATCGTTCCATCGTGGCAGCCACAGTACAAACAGCGCAGTGAGCTTGGTCCCAGG 540
Qy 541 GGAAGCCTCTACATCGGAGACGTGATCTCAATCCAGAGAAACCGGGGATCGATCTCCTT 600
Db 541 GGAAGCCTCTACATCGGAGACGTGATCTCAATCCAGAGAAACCGGGGATCGATCTCCTT 600
Qy 601 CAAAGGAACTAGTCTGATGGCGTCCCTCTGAAATCTGTCAGCCAGATGACACTTTGG 660
Db 601 CAAAGGAACTAGTCTGATGGCGTCCCTCTGAAATCTGTCAGCCAGATGACACTTTGG 660
Qy 661 GTGGAGAAATGCTTATGTTTGGGGGACGAATGTTAAACATTCAGATGTGCTTAGGGCGA 720
Db 661 GTGGAGAAATGCTTATGTTTGGGGGACGAATGTTAAACATTCAGATGTGCTTAGGGCGA 720
Qy 721 TTCTGTCGATTTCTCCACAATTAATCGTTTCGAGTGTCTCATGATCTTAAATCCAAAGTACATCC 780
Db 721 TTCTGTCGATTTCTCCACAATTAATCGTTTCGAGTGTCTCATGATCTTAAATCCAAAGTACATCC 780
Qy 781 AGATCATAGAGAGACTGTGGAGCGTGGAGGAGATCTCTAAATATCGACATGTGAGACA 840
Db 781 AGATCATAGAGAGACTGTGGAGCGTGGAGGAGATCTCTAAATATCGACATGTGAGACA 840
Qy 841 TTTATGACCATGATCTGATCTATAGCGGAAAATTTGTTTCGATACCCACTCGACATCATCC 900
Db 841 TTTATGACCATGATCTGATCTATAGCGGAAAATTTGTTTCGATACCCACTCGACATCATCC 900
Qy 901 CCTGTTTGACACTGAGTGTGAGGAATGCTACCTCTTTACTACCAAGTTTGAAGC 960
Db 901 CCTGTTTGACACTGAGTGTGAGGAATGCTACCTCTTTACTACCAAGTTTGAAGC 960
Qy 961 ATATTGAGCCAGACCTTTCAATCTCAAAAGCATCGGTGCACATGCGTGAACCTCAACCTTT 1020
Db 961 ATATTGAGCCAGACCTTTCAATCTCAAAAGCATCGGTGCACATGCGTGAACCTCAACCTTT 1020
Qy 1021 CAGATATAGACAAAATTTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
Db 1021 CAGATATAGACAAAATTTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
Qy 1081 CTGGAATTAAGGGGCTCTTCAAAATGTTAGTGTGCTCACTCGCCTCCGCTAGTTA 1140
Db 1081 CTGGAATTAAGGGGCTCTTCAAAATGTTAGTGTGCTCACTCGCCTCCGCTAGTTA 1140
Qy 1141 CAGTTGTTAAAGGGGGCTTTGAGGAGCCAAACAGGTGTGAAAAGCCAGAAATGTGACGAC 1200
Db 1141 CAGTTGTTAAAGGGGGCTTTGAGGAGCCAAACAGGTGTGAAAAGCCAGAAATGTGACGAC 1200
Qy 1201 GGAATGCTATGCTCTTATTCACATCGATGCACTTTTGCAAAATAGACAGATAGTCGTC 1260
Db 1201 GGAATGCTATGCTCTTATTCACATCGATGCACTTTTGCAAAATAGACAGATAGTCGTC 1260
Qy 1261 TTTCAAGAACTCCAGATGCCATTTCTGAAGGAGAGACTCCACACAGTCAGCATGTGT 1320
Db 1261 TTTCAAGAACTCCAGATGCCATTTCTGAAGGAGAGACTCCACACAGTCAGCATGTGT 1320
Qy 1321 TATACAACTATGTTGATGCTGTGAAGCCTTGAGATCGTATTGAGGTAAACAGGAGTTT 1380
Db 1321 TATACAACTATGTTGATGCTGTGAAGCCTTGAGATCGTATTGAGGTAAACAGGAGTTT 1380

Qy 1381 TCAAGGCCATGCGAGTTGCTCATGGCGCGCGACCTCACGTGAGGAATTCGACC 1440
Db 1381 TCAAGGCCATGCGAGTTGCTCATGGCGCGCGACCTCACGTGAGGAATTCGACC 1440
Qy 1441 TCCTTGTCTTCGCGACGCTTCCATCTTTTGTGTTTTAGTTTGCAGAAATCTTGATCGTGA 1500
Db 1441 TCCTTGTCTTCGCGACGCTTCCATCTTTTGTGTTTTAGTTTGCAGAAATCTTGATCGTGA 1500
Qy 1501 GTTGAGAAAAAGGGCGGTTGTTGTTGTTGAGGTGTTCTTGTGATGTTGTTGTCATCGAAAA 1560
Db 1501 GTTGAGAAAAAGGGCGGTTGTTGTTGTTGAGGTGTTCTTGTGATGTTGTTGTCATCGAAAA 1560
Qy 1561 ATAATGATGCACTTCGACATTTGAGCGCGTGTCTGCCCATATCTTTCGCAATCTGAAAGGAG 1620
Db 1561 ATAATGATGCACTTCGACATTTGAGCGCGTGTCTGCCCATATCTTTCGCAATCTGAAAGGAG 1620
Qy 1621 TGTCTACGCGCATTTGCCGCAAGTAAACATCACCGAGCTTCGACAAATGAGCCTCACCCGTGG 1680
Db 1621 TGTCTACGCGCATTTGCCGCAAGTAAACATCACCGAGCTTCGACAAATGAGCCTCACCCGTGG 1680
Qy 1681 CCGGGCGAGGGCCGTACGCGCAGACCCCTTACATCTGCGAGTTTCCGAGGAGGAGAGAAA 1740
Db 1681 CCGGGCGAGGGCCGTACGCGCAGACCCCTTACATCTGCGAGTTTCCGAGGAGGAGAGAAA 1740
Qy 1741 CGGATTCGCTCGTCTAGGAGGAGTCTCGAGTTTGGCAATCTCTGTTTATAGTT 1800
Db 1741 CGGATTCGCTCGTCTAGGAGGAGTCTCGAGTTTGGCAATCTCTGTTTATAGTT 1800
Qy 1801 CCCCTTACGATCGCGGGGACTCTCGGAACTCTCGGAACTCCAGTGGCTACTCCGGTTTACG 1860
Db 1801 CCCCTTACGATCGCGGGGACTCTCGGAACTCTCGGAACTCCAGTGGCTACTCCGGTTTACG 1860
Qy 1861 CTACCCCAGTCGTTACACTATGGGTACCCCATCGTTCCATCGTGGCAGCCACAGTACA 1920
Db 1861 CTACCCCAGTCGTTACACTATGGGTACCCCATCGTTCCATCGTGGCAGCCACAGTACA 1920
Qy 1921 AACAGCGCAGTGTGTTGTTTCCAGGGGAGCCCTTACATCGGAGACGTCGATCTCAAT 1980
Db 1921 AACAGCGCAGTGTGTTGTTTCCAGGGGAGCCCTTACATCGGAGACGTCGATCTCAAT 1980
Qy 1981 CCAGAGAACCCGGGCGATCGATCTCTCAAGGGAACCTAGTCTGATGGCGCTCTCTG 2040
Db 1981 CCAGAGAACCCGGGCGATCGATCTCTCTCAAGGGAACCTAGTCTGATGGCGCTCTCTG 2040
Qy 2041 AATCTGCTGAGCCAGATGACACTTTTGGGTGGAGAAATATGCTTATGTTTGGGGACGAAATG 2100
Db 2041 AATCTGCTGAGCCAGATGACACTTTTGGGTGGAGAAATATGCTTATGTTTGGGGACGAAATG 2100
Qy 2101 TTAACATTTCCAGATGTCTTAGGGCGATTCGTCGATTTCTCCACAAATATCGTTCAGTG 2160
Db 2101 TTAACATTTCCAGATGTCTTAGGGCGATTCGTCGATTTCTCCACAAATATCGTTCAGTG 2160
Qy 2161 CTCATGATCTTAATTTCCAAAGTACATCCAGATCATAGAGAGACTGTGGAGCGTGAGAGG 2220
Db 2161 CTCATGATCTTAATTTCCAAAGTACATCCAGATCATAGAGAGACTGTGGAGCGTGAGAGG 2220
Qy 2221 ATACTCTAAATATCGACATGTGAGACATTTATGACCATGATCTCTATATAGCGAAAAA 2280
Db 2221 ATACTCTAAATATCGACATGTGAGACATTTATGACCATGATCTCTATATAGCGAAAAA 2280
Qy 2281 TTGTTGATATCCCATCTGACATCTCCCTCTGTTGACATCTGAGTGTGAGGAATTTGCTA 2340
Db 2281 TTGTTGATATCCCATCTGACATCTCCCTCTGTTGACATCTGAGTGTGAGGAATTTGCTA 2340
Qy 2341 CCTCTTTACTACCAAGTTTGAAGAGCATATTTGAGGCCAGACTTTCAATCTCAAAAGCAT 2400
Db 2341 CCTCTTTACTACCAAGTTTGAAGAGCATATTTGAGGCCAGACTTTCAATCTCAAAAGCAT 2400
Qy 2401 CGGTGCACTATGCGTGAACCTCAACCTTTCAGATATAGACAAATTTGGTTCTTGTTTAAAGGAA 2460
Db 2401 CGGTGCACTATGCGTGAACCTCAACCTTTCAGATATAGACAAATTTGGTTCTTGTTTAAAGGAA 2460

QY 2461 TGGTTATCCGGTGCAGTCTATCATACCTGAAATTAAGGGGCCCTTCTTCAAATGCTTAG 2520
DB TGGTTATCCGGTGCAGTCTATCATACCTGAAATTAAGGGGCCCTTCTTCAAATGCTTAG 2520
QY 2521 TGTGTGGTCACTCGCTCCGCTAGTATTACAGTTGTTTAAAGGGCGGGTTCAGAGGCCAACAA 2580
DB TGTGTGGTCACTCGCTCCGCTAGTATTACAGTTGTTTAAAGGGCGGGTTCAGAGGCCAACAA 2580
QY 2581 GGTGTGAAAGCCAGAAATGTGCAGCACGGAATGCTATGTTCTTATTCAAAATCGATGCA 2640
DB GGTGTGAAAGCCAGAAATGTGCAGCACGGAATGCTATGTTCTTATTCAAAATCGATGCA 2640
QY 2641 CTTTGTGCAAAATGACAGATAGTGGCTCTTCAAGAACTCCAGATGCCATTCTCTGAAGGAG 2700
DB CTTTGTGCAAAATGACAGATAGTGGCTCTTCAAGAACTCCAGATGCCATTCTCTGAAGGAG 2700
QY 2701 AGACTCCACACACAGTCAAGTGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG 2760
DB AGACTCCACACACAGTCAAGTGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG 2760
QY 2761 GAGATCGTATTGAGGTAAACAGAGTGTTCAGAGCCATGCGAGTTCGAGTTGGTCCGAATC 2820
DB GAGATCGTATTGAGGTAAACAGAGTGTTCAGAGCCATGCGAGTTCGAGTTGGTCCGAATC 2820
QY 2821 AACGAACATTACGAGTGTGTTATAGACCTACATCGATTGCGTCAAGAGTCTG 2880
DB AACGAACATTACGAGTGTGTTATAGACCTACATCGATTGCGTCAAGAGTCTG 2880
QY 2881 ACAGGGTTCGACTGCAAACTGAAGTCTTATGAGATGAGTAAAGAGATGATATGATG 2940
DB ACAGGGTTCGACTGCAAACTGAAGTCTTATGAGATGAGTAAAGAGATGATATGATG 2940
QY 2941 CTGGGTATCATGAAAGTGTATCTCAGAGCTGCTAAATGAAGCAAGATTCAAAAACCTTA 3000
DB CTGGGTATCATGAAAGTGTATCTCAGAGCTGCTAAATGAAGCAAGATTCAAAAACCTTA 3000
QY 3001 AAGAGCTGTCAGCTCCGGGCACTTATGATAGACTTTTCAAGTGCCTGCGTCCAAAGCA 3060
DB AAGAGCTGTCAGCTCCGGGCACTTATGATAGACTTTTCAAGTGCCTGCGTCCAAAGCA 3060
QY 3061 TTTGGAGCTCAAGATATTAAGAGGTCTCTTTCAGAGCTCTTTCAGAGTCTTTCAGAGTCT 3120
DB TTTGGAGCTCAAGATATTAAGAGGTCTCTTTCAGAGCTCTTTCAGAGTCTTTCAGAGTCT 3120
QY 3121 AGAAAATTCATCTGAGCATCTTTCGAGGTGACATCAATGTTTACTTGTTCGGGACC 3180
DB AGAAAATTCATCTGAGCATCTTTCGAGGTGACATCAATGTTTACTTGTTCGGGACC 3180
QY 3181 CTGGTACCAAGTAAATCTCAGCTGCTTCAATATGTCACAGATAGCTCTCTCGTGAATCT 3240
DB CTGGTACCAAGTAAATCTCAGCTGCTTCAATATGTCACAGATAGCTCTCTCGTGAATCT 3240
QY 3241 ACATAGTGGGCGAGAAATGTCGGCGTGGGTGACAGGTATGTAACGAAGATCCAG 3300
DB ACATAGTGGGCGAGAAATGTCGGCGTGGGTGACAGGTATGTAACGAAGATCCAG 3300
QY 3301 AACTCGAGAGCGTATTGAGAGCGGAGCTTGGTCTTCTAGTATCGTGGGATAGCT 3360
DB AACTCGAGAGCGTATTGAGAGCGGAGCTTGGTCTTCTAGTATCGTGGGATAGCT 3360
QY 3361 GTATCGATGAGTTCGACAAAATGTCGATAATGCCGAAGCATCTTCATGAGTAAATGG 3420
DB GTATCGATGAGTTCGACAAAATGTCGATAATGCCGAAGCATCTTCATGAGTAAATGG 3420
QY 3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTCCTCGCTGAACCGTCCGACCT 3480
DB AGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTCCTCGCTGAACCGTCCGACCT 3480
QY 3481 CTGTCTTCGATGTGCAAAATCTAGTGGGTCCGATACAAATGCGCGCCCTTCTGTGATG 3540
DB CTGTCTTCGATGTGCAAAATCTAGTGGGTCCGATACAAATGCGCGCCCTTCTGTGATG 3540
QY 3541 ATAAACATCCAGCTTCTCCCAACTCTACTTCTTAGATTTGATTTAAATTTAAATGCTCG 3600

DB ATAAACATCCAGCTTCTCCCAACTCTACTTCTTAGATTTGATTTAAATTTAAATGCTCG 3600
QY 3601 ACAAAACAGAGCAGCAAAACCGATCGTCTCGCCAGGACATCTCGTGGCTTTTACACTATG 3660
DB ACAAAACAGAGCAGCAAAACCGATCGTCTCGCCAGGCACTCTCGTGGCTTTTACACTATG 3660
QY 3661 AAAACTATGAAGTTTCAAAGCAGGCGCTTAGATCTCAAAACACTTACCGGCTATATCA 3720
DB AAAACTATGAAGTTTCAAAGCAGGCGCTTAGATCTCAAAACACTTACCGGCTATATCA 3720
QY 3721 CCTATGCTCGTCAGCATGTACATCTATTAAGTGAAGTCTCTGAAGATTTGATTA 3780
DB CCTATGCTCGTCAGCATGTACATCTATTAAGTGAAGTCTCTGAAGATTTGATTA 3780
QY 3781 ATGGCTATGTTGAGATGCGCCAAAGGGCAACTTCTTCTGGAAGCAGTAAAGGTGATAA 3840
DB ATGGCTATGTTGAGATGCGCCAAAGGGCAACTTCTTCTGGAAGCAGTAAAGGTGATAA 3840
QY 3841 CAGCCACACTCGGCAACTCGAAAGTATGATTCGATATCAGTGAAGCCCTAGCTCGAATGA 3900
DB CAGCCACACTCGGCAACTCGAAAGTATGATTCGATATCAGTGAAGCCCTAGCTCGAATGA 3900
QY 3901 GATTTCTGAAGTGTAGAGAAAGTTGATGAGCAGAGAGCTGTGCGCCTTTTAGAGCTCG 3960
DB GATTTCTGAAGTGTAGAGAAAGTTGATGAGCAGAGAGCTGTGCGCCTTTTAGAGCTCG 3960
QY 3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGATCTTTATCACA 4020
DB CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGATCTTTATCACA 4020
QY 4021 CTGGAGTGTGCGCCAGCAGCGGTATTCGTGCGGCCAACTTGTAGCTCTGCGAGAGC 4080
DB CTGGAGTGTGCGCCAGCAGCGGTATTCGTGCGGCCAACTTGTAGCTCTGCGAGAGC 4080
QY 4081 TTATAGCAGATAAAATTTTCACTGCGAGCTCTCTGCTTGAAGCAGCTCAGCTCTTGG 4140
DB TTATAGCAGATAAAATTTTCACTGCGAGCTCTCTGCTTGAAGCAGCTCAGCTCTTGG 4140
QY 4141 AGGATATCCGAGCCAAAGCAGGTGAGTGTAGTTTGCAGGATATTAATAATGCTCTCG 4200
DB AGGATATCCGAGCCAAAGCAGGTGAGTGTAGTTTGCAGGATATTAATAATGCTCTCG 4200
QY 4201 GTAGCTCCAAAGGAGAAGCTTTCTTACTGTCCATGTTGACATAGTCAAGAGATTTGAG 4260
DB GTAGCTCCAAAGGAGAAGCTTTCTTACTGTCCATGTTGACATAGTCAAGAGATTTGAG 4260
QY 4261 ACAGTTTCTAACTGTTTGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320
DB ACAGTTTCTAACTGTTTGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320
QY 4321 TTTCCCATGCGCAATTTCCAGAGCTCGC 4348
DB TTTCCCATGCGCAATTTCCAGAGCTCGC 4348

RESULT 2

US-09-828-062-2
; Sequence 2, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNET, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/09/828,062
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07

; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 723
 ; TYPE: DNA
 ; ORGANISM: Physcomitrella patens
 US-09-828-062-2

Query Match 15.5%; Score 673; DB 3; Length 723;
 Best Local Similarity 97.1%; Pred. No. 2.2e-209;
 Matches 696; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

Qy	2753	GAAGCTCGAGATCGTATTAGGTAACAGAGGTTTCAAGGCCATGGCAGTTCAGTTGG	2812
Db	3	GCACAGGAGATCGTATTAGGTAACAGAGGTTTCAAGGCCATGGCAGTTCAGTTGG	62
Qy	2813	TCCGAATCAACGAACTTACGAGCATTTGTAAGACCTTACATCGATTCCGTCAGCTCAA	2872
Db	63	TCCGAATCAACGAACTTACGAGCATTTGTAAGACCTTACATCGATTCCGTCAGCTCAA	122
Qy	2873	GAAATCTCACAGGGTCGACTGCAAACTCAAGATCCTATGGAGATGGATAAGGAAATGA	2932
Db	123	GAAATCTCACAGGGTCGACTGCAAACTCAAGATCCTATGGAGATGGATAAGGAAATGA	182
Qy	2933	TATGATGCTGGGTATCATGAAGTGATATTCAGAGCTGCTATGAAGCAAAAGATTCA	2992
Db	183	TATGATGCTGGGTATCATGAAGTGATATTCAGAGCTGCTATGAAGCAAAAGATTCA	242
Qy	2993	AAACTTAAAGCTGCTCCAGCTCCGGGCAATTTATGATGACATTCAGAGTCCGCTGGC	3052
Db	243	AAACTTAAAGCTGCTCCAGCTCCGGGCAATTTATGATGACATTCAGAGTCCGCTGGC	302
Qy	3053	TCCAAGCAATTTGGGAGCTTGAAGATATTAAAGAGGTCTCTTTGCCAGCTCTTTGGTGG	3112
Db	303	TCCAAGCAATTTGGGAGCTTGAAGATATTAAAGAGGTCTCTTTGCCAGCTCTTTGGTGG	362
Qy	3113	GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTGT	3172
Db	363	GAAGCTTAAGAAATTTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTGT	422
Qy	3173	TGGGACCCCTGTACAGTAATCTCAGTGTCTTCAAGTATGGCAAGATAGCTCCCTCG	3232
Db	423	TGGGACCCCTGTGTACAGTAATCTCAGTGTCTTCAAGTATGGCAAGATAGCTCCCTCG	482
Qy	3233	TGGAATCTACACTAGTGGCGGAGGAGTTTCGGCGTTGGGCTGACAGCGTATGT-AAACA	3291
Db	483	TGGAATCTACACTAGTGGCGGAGGAGTTTCGGCGTTGGGCTGACAGCGTATGTAAACA	542
Qy	3292	AGGATCCAGAACTCGAGAGACGGTATTCGAGAGCGGAGCTTTGGTTCCTTAGTATCGTG	3351
Db	543	AGGATCCAGAACTCGAGAGACGGTATTCGAGAGCGGAGCTTTGGTTCCTTAGTATCGTG	602
Qy	3352	GGATATGCTATCGATGAGTTTCGACAAATGCTGATTAATGCCCGAAGCATGCTTCATG	3411
Db	603	GGATATGCTATCGATGAGTTTCGACAAATGCTGATTAATGCCCGAAGCATGCTTCATG	662
Qy	3412	AGTATGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTTGCTCGCTGA	3468
Db	663	AGTATGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTTGCTCGTGAAGCCGA	719

RESULT 3

US-09-604-605-1
 ; Sequence 1, Application US/09604605
 ; Patent No. 6421613
 ; GENERAL INFORMATION:
 ; APPLICANT: Ramgopal Nadimpalli
 ; APPLICANT: Carl R. Simmons
 ; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
 ; FILE REFERENCE: 1138
 ; CURRENT APPLICATION NUMBER: US/09/604,605
 ; CURRENT FILING DATE: 2000-06-27
 ; PRIOR APPLICATION NUMBER: 60/143,222

; PRIOR FILING DATE: 1999-07-09
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2533
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (110)...(2269)
 US-09-604-605-1

Query Match 4.8%; Score 209.6; DB 3; Length 2533;
 Best Local Similarity 52.0%; Pred. No. 4.2e-57;
 Matches 498; Conservative 0; Mismatches 454; Indels 6; Gaps 1;

Qy	2992	AAAACTTAAAGAGCTGTCCAAAGCTCCGGGCAATTTATGATAGACTTTTCAAGGTGCGTGG	3051
Db	1059	AACAATTTGACCGATTGGCTGAGGACGGTGATATCTACAGTAAGTGGCAAGTCCCTGG	1118
Qy	3052	CTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGTCTTTCTTTGCCAGCTCTTTGGTG	3111
Db	1119	CACCTGAAATATTATTGGCCATGAAGATGTCAAAAAAGGCACCTGCTGTACTTCTTTGGTG	1178
Qy	3112	GGAAAGGCTAAGAAATTTCCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTTACTTG	3171
Db	1179	CACCCATTCGGAAGCTCGCAGATGGCAATGAAGATCAGAGGAGACCTGCAATATATGATGA	1238
Qy	3172	TTGGGGACCCCTGGTACAGTAAATCTCAGCTGCTTCAAGTATGTGCAAGATAGCTCCTC	3231
Db	1239	TGGAGATCTCTGGTGTGCAAAAGAGTCAACTTCTGAGCATATTATCAATGTTGCCACAA	1298
Qy	3232	GTGGAATCTACACTAGTGGGCGAGAAATTCGGCGGTTGGGCTGACAGCGTATGTAAACA	3291
Db	1299	GAGGAGTGTACACCACTGGACGTGGAGCAGTGGTGTGGACTTACCCTGCAAGTCCAGA	1358
Qy	3292	AGGATCCAGAACTCGAGAGACGGTATTTGGAGAGCGGAGCTTTGGTCTTCTAGTATCGTG	3351
Db	1359	AAGATCCAGTAAACAAATGAGTTTGTCTCGAAGGTGGAGCACTGGTACTTGGCAGATATGG	1418
Qy	3352	GGATATGCTGTATCGATCAGTTTCGACAAATATGCTGATAATGCCCGAAGCATGCTTCATG	3411
Db	1419	GCATTTGTGCTATAGATGAGTTTGACAGATGGAGAGTTCAGACAGGACAGCGATTCTATG	1478
Qy	3412	AGTAAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCAATGCTCGCTGAAGC	3471
Db	1479	AGTAAATGGAGCAACAAACGGTATCTGTAGCCAAAGGTGGCATCACCACTCTCTTAATG	1538
Qy	3472	CTCGAGCGTCTGTCTTCATGTGCAAAATCCTAGTGGGTCCCGATACAAATGCGCGCTTT	3531
Db	1539	CGAGAACTGCAATTTCTGGCTGCTGCAAAATCCAGCATGGGGAAGGTATGATATGAGAGAA	1598
Qy	3532	CTGTGATTTGATAAATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTTACT	3591
Db	1599	CCCCAGCAGAAATATAAATCTACCTCAGCTCTTCTGCTGCTTTCGACCTCTTGTGT	1658
Qy	3592	TAATGCTCGACAAACAGACGAGCAAAACGATCGTGTCTCCCGAGGCAATCTCGTGCTT	3651
Db	1659	TAATCTCGATCGTCAGACATGGAAACTGATCTTTGAAATGCAAGACACGTTGTTCTATG	1718
Qy	3652	TACACTATGAAACCTATGAAGTTTCAAAGCAGGACGCTTAGATCTACAAACACTTACCG	3711
Db	1719	TGCATCAAAATCTTGAATCACCAGCGCTGGGGTTTCAACACTTGAAGCATCTGTACTCA	1778
Qy	3712	CGTATATCACCTATGCTCGTACGATGTACATCTTACATTAAGTGTAGAGCTGCTGAAG	3771
Db	1779	GAGCATAACATCTGCTGCTAGAAAGTCAATCTCTTCTGTTCTTAGAGAGCTCGAGAAAT	1838
Qy	3772	ATTGATTAATGGCTATGTTAGATGCGCAAAAGGGCAACTTTCTCTGGAAGCAGTAAAA	3831
Db	1839	ACATTGCAACTGCATATTCCAGCATCCGCGCAAGAGGAGGCGAAGTCAAAATGCACCAACCT	1898
Qy	3832	AGGTGATACAGCCACACCTCCGGCAACTCGGAAAGTATGATTCGTATCAGTGAAGCCCTAG	3891

Db 1899 CCTACAACTATCAGAAC-----ACTTTTGAGCATACTCCGGTATCTCTATTGCGCTTGG 1952
QY 3892 CTCGAATGATGATTTCTGAAAGTGGTAGAGAAAGTTGATGACGACGAGAGCTGTGCGCT 3949
Db 1953 CAAGACTGAGGTCTTCAGAAACTGTGGCTCAGAGCGATGTGAGGAAGCACTGGGACT 2010

RESULT 4

US-10-101-080-1
; Sequence 1, Application US/10101080
; Patent No. 6879916
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/10/101,080
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/604,605
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110)...(2269)
US-10-101-080-1

Query Match 4.8%; Score 209.6; DB 3; Length 2533;
Best Local Similarity 52.0%; Pred. No. 4.2e-57;
Matches 498; Conservative 0; Mismatches 454; Indels 6; Gaps 1;
QY 2992 AAAAATTAAAGAGCTGTCCAGCTCCGGGCAATTATGATAGACTTTCAGGTGCTGG 3051
Db 1059 AACAAATTGACGATTTGGCTGAGGAGCGTGATATCTACAGTAAGCTGGCAAGTCTCTGG 1118
QY 3052 CTCGAGCATTTGGGAGCTTGAGATATTAAAGGGTCTTCTTCCAGCTTTTGGTG 3111
Db 1119 CACCTGAATATTGCGCCATGAAGATGTCAAAAGGCACTGCTTTACTTACTTGTGGTG 1178
QY 3112 GGAAGGCTAAGAAAATTCCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTACTTG 3171
Db 1179 CACCCCATCGAAGCTCGCAGATGGCATGAAGATCAGAGGACCTGCATATATGATGA 1238
QY 3172 TTGGGACCTGGTACAGTAATCTCAGCTGCTTCAGTATGTGCAAGATAGCTCCTC 3231
Db 1239 TGGGAGATCTGTGTTCGAAAGAGTCACTTCTGAAGCATATTATCAATGTTGCACCAA 1298
QY 3232 GTGGAATCTACATAGTGGGAGGAGGTTCCGGGTTGGGCTGAGCAGGTATGTAAACGA 3291
Db 1299 GAGGAGTGATACCACTGACCTGGGAGCAGTGTGTGGACTTACCGCTGCGAGTCCAGA 1358
QY 3292 AGGATCCAGAACTCCGAGACGGTATTCGAGAGCGGCTTTGGTCTTCTAGTGATCGTG 3351
Db 1359 AGATCCAGTAACAAATAGTTTGTCTCGAAGTGGAGCACTGGTACTGGCAGATATGG 1418
QY 3352 GGATATGCTGTATCGATAGTTTCGACAAAATGTCTGATATGCCGGAAGCATGCTTTCATG 3411
Db 1419 GCATTTGTCTATAGATAGTTTGACAAAGTGAAGATCAGACAGCAGCGAATTCATG 1478
QY 3412 AGTAAATGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTTGCCCTCGCTGAACG 3471
Db 1479 AGGTAATGAGCAGCAACAGATTAGCATTTGCCAAGCTGGCATCACCACTCTCTTAAATG 1538
QY 3472 CTGCGAGCTGTGCTTGTGATGTGCAAACTCTAGTGGTCCCGATCAATGCGCGCTTT 3531
Db 1539 CGAAGACTGCAATCTCTGGCTGCTGCAAACTCCAGCATGGGGAAGGTATGATATGAGGAGAA 1598
QY 3532 CTGTGATTGATACATCCAGCTTCTCTCAACTCTACTTCTAGATTGATTTAATTTACT 3591

Db 1599 CCCGAGCAGAAAATATAAATCTACCTCCAGCTCTTCTGTCTCGTTTCGAGCCTCTTTGGT 1658
QY 3592 TAATGCTCGACAAAACAGACGAGCAAAACGATCGTCTCGCAGGCACTCTCGTGGCTT 3651
Db 1659 TAATCTCGATCGTCGAGACATGGAACCTGATCTTGAATGGAAGACACACGTTGTTCA 1718
QY 3652 TACACTATGAAAACCTATGAAAGTTTCAAGCAGAGACGCTTATGATCTTACAAACACTTACCG 3711
Db 1719 TGCATCAAAATCTTGAATCACCGCGCTGGGGTTACACCACCTTGAGCCATCTGTACTCA 1778
QY 3712 CGTATATCACCTATGCTCGTCAGCATGTACATCCACATTAAGTATGAGCTGCTGAAG 3771
Db 1779 GAGCATACATATCTGCTGTCTAGAGAGTCAATCTCTCTGTCTCTAGAGAGCTCGAGAA 1838
QY 3772 ATTTGATTAATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCCTGGAAGCAGTAATA 3831
Db 1839 ACATTCGCACTGCATATTCAGCATCCGCAAGAGGAGGCGAAGTCAATGCAACCACT 1898
QY 3832 AGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAG 3891
Db 1899 CCTACACAACTATCAGAAC-----ACTTTTGAGCATACTCCGTATCTCTATTGCTTGG 1952
QY 3892 CTCGATGAGATTTCTGAGTGGTAGAGAAAGTTGATCGACGACGAGCTGTGCGCT 3949
Db 1953 CAAGACTGAGGTCTCAGAAACTGTGGCTCAGAGCGATGTGCAAGCACTCGGACT 2010

RESULT 5

US-09-248-796A-4534
; Sequence 4534, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4534
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4534

Query Match 4.6%; Score 198.6; DB 3; Length 1635;
Best Local Similarity 55.8%; Pred. No. 1.2e-53;
Matches 378; Conservative 0; Mismatches 299; Indels 0; Gaps 0;
QY 2979 GAAGCAAGATTCAAAACTTAAAGAGCTGTCCAGCTCCCGGGCAATTTATGATAGACTT 3038
Db 445 GAAGAGAAATACGTGAATTTAGAAAATATCACACGAAAAAGGTATTTATGATAAAATC 504
QY 3039 TCAGGTCCTGGCTCCAAAGCAATTTGGGAGCTTTGAAGATATTAAAAAGGCTTCTTTGCG 3098
Db 505 ATTGCCCTCAATGGCTCCATCCATATATGCTCAATAAAGATATTAAAACTGCAATGCTGT 564
QY 3099 CAGCTCTTTGGTGGGAGGCTAAGAAAATTCATCTGGAGCATTTTCCGAGGTGACATC 3158
Db 565 TCATTTATTTGGTGGTGTGCTTAAAGATGTTAATGGGAAACTTTTCCATTTAGAGTGATTT 624
QY 3159 AATGTTTACTTGTGTTGGGACCTCGTGTACAGTAAATCTCAGCTGCTTCAGTATGTGCAC 3218
Db 625 AATGACTATTATTAGGGGATCCAGGTACTGCTTAATCGAAATATTGAAATATGCTGAG 684
QY 3219 AAGATAGCTCTCGTGGAAATCTACACTAGTGGCGGAGGAGTTTCGGCGGTGGGCTGACA 3278
Db 685 AAAACTGCTAGTCTGCTGTTTTTGGCAACTGGGTCAAGGTGCGCTCGGCTGTTGCTTTACA 744

QY 3279 GCGTATGTAAAGGATCCAGAACTCGAGAGCGGTATTGGAGAGCGGAGCTTTGGTT 3338
Db 745 GCATCGGTTAGAAAAGACCAATCACCTCGTGAATGGACATTAACAAGGTGGGCTTTGGTG 804
QY 3339 CTTAGTGATCGTGGATATGCTGATCGATGAGTTCGACAAATGCTGTGATAATGCCGA 3398
Db 805 CTTGCAGATAAAGGAGCTTGTCTTATTGATGAATTTGATAAGATGAATCAAGATAGA 864
QY 3399 AGCATGCTTCATGAGTAAATGAGGACCAACAAACGGTATCTGTAGCCAAAGGGGTATCAAT 3458
Db 865 ACTTCTATTATGAGCTATGAAACAAAGATATCTATTCTTAAAGCTGGTATCGTC 924
QY 3459 GCCTCGCTGAACGCTCGGACGCTCTGCTTCGTATGTCAGTGCATAATCCCTAGTGGGTCGCGATAC 3518
Db 925 ACTACCTTACATGCCAGATGCTGCTTATCGTCTGCCAATCTTAATGGAGGTCGATAC 984
QY 3519 AATGGCGGCTTTCTGTGATGATTAACATCCAGCTTCTCCAACTCTACTTTCTAGATTT 3578
Db 985 AACTCTACATTACCAATTACCAGAAATGTCGATTTAACTTACCCAAATTTTATCCAGATTT 1044
QY 3579 GATTTAAATTTACTTAATGCTCGACAAACAGACGAGCAAAAGATCGTCTCGCCAGG 3638
Db 1045 GATATTATGTGATCGTGGAGATTTGGTCAATCTGATCTGATGAACGATTAGCTAGT 1104
QY 3639 CATCTCGTGGCTTTACA 3655
Db 1105 TTTGTCAATGATCCCA 1121
RESULT 6
US-09-248-796A-5904
; Sequence 5904, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5904
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (2)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno
US-09-248-796A-5904

Query Match 4.5%; Score 197.6; DB 3; Length 1488;
Best Local Similarity 51.3%; Pred. No. 2.5e-53;
Matches 514; Conservative 0; Mismatches 479; Indels 9; Gaps 2;
QY 2939 TGCTGGGTATCATCAAGATGATACCTTCAGAAAGCTGCTTAATGAAGCAAGATTCAGAAACT 2998
Db 144 TGATTTGAATGACCAAGAGTTTCTTCTACATCTTTGAGTGATGCTGAGTATTAACAATT 203
QY 2999 TAAAGAGCTGCCAAGCTCCCGGCAATTTATGATAGATTTCAAGTCCGCTGCTCCAAG 3058
Db 204 AAAGGAAATGGTCAAGGATGAACATATTTATGACAAAGCTTGTGAACCTCAATTGCCACGC 263
QY 3059 CATTTGGGAGCTTGACATATTAAGAGGTCTCTTTGTCAGCTCTTTGGTGGGAAGGC 3118
Db 264 TGTGTTGGCCACGAGTGATCAAAAGGAATCTTTTTCGAATTTATGCGCGGTGTCA 323
QY 3119 TAAGAAAAATCCATCTCGGAGCATCTTCCGAGGTGACATCAATGTTTCTTGTGGGA 3178

Db 324 CAAACAAAGTAGATGAATCAATTAAGAGGGATATAAATTTGTTGTTGTA 383
QY 3179 CCCTGGTACAGTAAATCTCAGCTGCTTATGATGTGCAACAAGATAGCTCTCTGGAAT 3238
Db 384 CCCATCCACTTCGAAATCACAGTTTTTGAAGTATGCTGTGGGTTTTTCGCTCGTGT 443
QY 3239 CTACACTAGTGGGAGGAAGTTCCGGGTTGGCTGACAGCGTATGTAAACGAAGGATCC 3298
Db 444 CTACACTTCTGGTAAGGCTTCTCTGCTGCTGGGTTGACAGCGCTGTAGTGAAGGACGA 503
QY 3299 AGAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTCTTCTAGTGATCGTGGGATG 3358
Db 504 AGAAGCGGAGATACACTATTGAAGCAGGTGCTTTGATGTTGCGAGATAACGGTATTG 563
QY 3359 CTGTATCGATGAGTTCGACAAAATGCTGTGATAATGCGCGGAAGCATGCTTCAATGAGTAA 3418
Db 564 TGCAATTTGATGAATTTGACAAAGATGATATTGCTGACCAAGTTGCTATTTCAGAACCCAT 623
QY 3419 GGACCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTCGCTCGCTGGAACGCTCGAC 3478
Db 624 GGAACAAACAAATCTCCATTGCAAAAGCTGGTATTTCATCTACTCTTAAATGCCCGTAC 683
QY 3479 GTCTGTCTCTGCATGTGCAAAATCCTAGTGGGTCCGATACAATGCGCGCTTTCTGTGAT 3538
Db 684 ATCTATCTTAGCAGCTGCTAACCTTATTTGGGGTAGATATAATAGAAAATTTGGGTCTTCG 743
QY 3539 TGATAACATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTTACTTAATGCT 3598
Db 744 TTCTAAATTTGAATATGACAGCACCAATTTATGTCGAGATTTGATTTGTTGTTGTTCT 803
QY 3599 CGACAAACAGACGAGCAAAAGATCGTCTGCTCGCAGGATCTCGGTGCTTTACACTA 3658
Db 804 TGATGATTTGAATGAAAGATTTGATCTCAATTAGCATCATCACACATGTTGATTGACAT 863
QY 3659 TGAAACTACTGAAGTTTCAAAGCAGGACGCTTAGATCTACAACACATTTACCGCGCTATAT 3718
Db 864 GCTTAGACAGCATGCAATTTGATCTACCTTACTCAGCCCAACAAATTA-----GCAAGATA 917
QY 3719 CACCTATGCTCGTCAGCATGTACATCTTACATTAAGTATGATGAGTCTGGAAGATTTGAT 3778
Db 918 CATTAATAACGCAAGAGCTTTCAAGCTTAGAATGATTAAGAAAGCTAGAGACTTTTGTGT 977
QY 3779 TAATGGCTATGTTGAGATGCGCAAAAGGGCAACTTTCTCGAAGCAGTAAAAAGTGAT 3838
Db 978 CACTAGATACAAGGAATTTGAGAGAAAGATG---ATGCTCAAGGTTTAGGGAGGTCTCTCGTA 1034
QY 3839 AACAGCCACCTCGGCAACTCGAAAGTATGATTTGATCTAGTGAAGCCCTTAGCTCGAAT 3898
Db 1035 CAGAATTTACTGTTAGACAAATTTGGAGTCCATGATTAATGTTGAGAAAGCTATAGCCAGAC 1094
QY 3899 GAGATTTTCTGAAAGTGGTAGAGAAAGTTGATCGACGAGAACG 3940
Db 1095 CAATTGTACCGAAGAAATAACGCCAAGTTTGTGCTGAGGC 1136
RESULT 7
US-09-949-016-2468
; Sequence 2468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2468
; LENGTH: 2917
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-2468

Query Match 4.3%; Score 186.4; DB 3; Length 2917;
Best Local Similarity 51.7%; Pred. No. 1.9e-49;
Matches 477; Conservative 0; Mismatches 436; Indels 9; Gaps 2;
QY 3025 TTTATGATAGACTTTCAAGGTCGCTCGCTCCAGCATTTGGGAGCTTTGAAGATATTAATAA 3084
DB 1098 TATACCACAATCTTTGTACCAGCTGTTCCTACTATACATGCGCAATGATGAAGTAAAC 1157
QY 3085 AGGCTCTTTTCCAGCTCTTTGGTGGGAAGGCTTAAGAAAATTCATCTGGAGCATCTT 3144
DB 1158 GGGGTGTCTGCTGATGCTCTTTGGTGGGCTTCCAAAGACAACAGGGAAGGACCTCTC 1217
QY 3145 TCCGAGTGCATCAATGTTTACTTTTGGGAGCCCTGCTACCACTAAATCTCAGCTGC 3204
DB 1218 TTGAGGGGACATAAATGTTTGCAATTTGGTGACCCCAAGTACAGCTAAGAGCCAATTC 1277
QY 3205 TTCAGTATGTGCAAGATAGCTCTCTGCGGAATCTACACTAGTGGCGGAGGAAGTTCG 3264
DB 1278 TCAGACAGCTGGAGGAGTTCAGCCCCAGAGCTGTCTACACAGTGTGAAGCGTCAGTG 1337
QY 3265 CGGTGGGCTGACAGCGTATGTAAACGAAGGATCCAGAAATCCGAGAGCGGTATTGGAGA 3324
DB 1338 CTGCTGGCTTAACAGCAGCTGTTGTGAGAGATGAAGAACTCATGAGTTTGTTCATTGAG 1397
QY 3325 GGGAGCTTTGGTCTTATGATGCTGGGATATGCTGTATCGATGATGTTGCAAAATGT 3384
DB 1398 CTGAGCTTTGATGTTGGCTGATAATGCTGTGTTGTTGATGATGAATTTGATAAGATG 1457
QY 3385 CTGATAATCCCGAAGCATGCTTCATGAGTATGAGCAACAAACGGTATCTGTAGCCA 3444
DB 1458 ACGTGGGATCAAGTTGCTATTCATGAAGCTATGAAACAGACAGACATATCCATCACTA 1517
QY 3445 AAGGGGTATCATTTGCTCGCTGAACGCTCGGAGCTGTCTCTGCTGCTATGCAAAATCCTA 3504
DB 1518 AAGCAGAGTGAAGGCTACTCTGAACGCCGAGCTCCATTTTGGCAGCAGCAACCCAA 1577
QY 3505 GTGGTCCGATACAATCGCGCTTTCTGTGATGATGAATCAATCCAGCTTCTCCAACTC 3564
DB 1578 TCAGTGCACATATGACAGATCAAAATCATTTGAACAGAAATATAAATTTGTCACTCC 1637
QY 3565 TACTTTCTAGATTGATTTTAATTTTACTTAATGCTCGACAAACAGAGGAGCAAAACGATC 3624
DB 1638 TCATGTCCCGATTCGATCTCTCTTTATCTTGTGGATGAATGTAATGAGGTTTACAGATT 1697
QY 3625 GTGCTCTCCCGGATCTCGTGGCTTTACACTATGAAACTATGAAGTTTCAAGCAGG 3684
DB 1698 ATGCCATGCGCGCGATAGATGATTCG- ---ATTCAAGAAATGAGGAATCAATTTGATC 1754
QY 3685 ACCGCTTAGATCTACAAACACTTACCGGCTATATCACCTATGCTCGCTCAGCATGTATC 3744
DB 1755 GTGCTATTTCCCTCGATGATATCAGAAGATATCTTCTTTTGAACAGAGTTTAAC- --- 1811
QY 3745 CTACATTTAAGTGAAGCTGCTGAAGATTGATTAATGCTATGTTGAGATCGGCCAA 3804
DB 1812 ---CCAAGATTTCCAAGAGTCAAGAGCTTCATTTGGAGCAATATAACATCTCCGCC 1868
QY 3805 AGGGCAACTTCTCGAGAGCAGTAAAGGTGATACAGGCCACCTCGGCACTCGAAA 3864
DB 1869 AGAGAGATGTTCTGGAGTGACCAAGTCTTTCATGAGGATTTACAGTGCAGCAGCTTGA 1928
QY 3865 GTATGATTCGTATCAGTGAAGCCTAGCTCGATGAGATTTTCTGAGTGGTAGAGAAAG 3924
DB 1929 GCATGATTCGTCTCTGAAGCTATGGCTCGGATGCACTGCTGTGATGAGGTCCAACCTA 1988
QY 3925 TTGATGACGAGAAAGCTGTGCG 3946

DB 1989 AACATGTGAAGGAGCTTTCCG 2010

RESULT 8
US-09-949-016-5021
; Sequence 5021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5021
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-5021

Query Match 4.2%; Score 183.8; DB 3; Length 2533;
Best Local Similarity 56.2%; Pred. No. 1.2e-48;
Matches 366; Conservative 0; Mismatches 282; Indels 3; Gaps 1;
QY 3006 CTGTCCAAGCTCCGGGCATTTATGATAGACTTTCAAGTCTGCTGCTCAAGCATTTGG 3065
DB 1055 CTGGCTGCCCTCCCAATGCTATGAGGTCACTCCAAAGAGCATCGCCCCCTCATCTTT 1114
QY 3066 GAGCTTGAAGATATTAAGAGGCTCTTTCGCCAGCTCTTTGGTGGGAAGGCTAAGAA 3125
DB 1115 GGGGCGACAGACATGAAGAGGCCATTTGCTGCTGCTCTTTGGGGGCTCCCGAAGAGG 1174
QY 3126 ATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGTGGGACCTCGT 3185
DB 1175 CTCCCTGATGAGTACTCTCCGAGGAGACATCACTGCTGATGCTAGGGGACCTGGG 1234
QY 3186 ACCAGTAAATCTCAGCTGCTTCAGTATGTGCAAGATAGCTCTCTCGTGAATCTACAT 3245
DB 1235 ACAGCCAACTCCAGCTCTTGAAGTTTGTGGAGAGTGTCTCCCATTTGGGGTATACAG 1294
QY 3246 AGTGGCGAGGAAGTTTCGGCGGTGGGCTGACAGCGTATGTAAAGAGGATCCAGAACT 3305
DB 1295 TCTGGGAAAGGCGAGCAGCGAGCTGAGCTGAGAGCTCGGTGATGAGGGACCTTCGTC 1354
QY 3306 CGAGAGACGATTTGGAGAGCGGAGCTTTGGTCTTTAGTATCGTGGGATGCTGTATC 3365
DB 1355 CGGAATTTTCATGAGGGCGAGCCATGGTCTCGCGATGGTGGGGTCTGTGTATT 1414
QY 3366 GATGAGTTCGACAAAATGCTGATAATGCCGAAGCATGCTTCATGAGGTAATGAGGCAA 3425
DB 1415 GACGAGTTTGAACAGATGCGAGAGATGACCGTGTGGCAATCCACGAGCCATGAGCAG 1474
QY 3426 CAAACGGTATCTGTAGCCAAAGGGGTATCATTTGCTCGCTGAGAACGCTCGGACGTCTGC 3485
DB 1475 CAGACCATCTCTATGCCAAAGGCTGGGATCACCACCACTGAACTCCGCTGCTCGTC 1534
QY 3486 CTTGATGTGCAAAATCCTAGTGGGTCGCGATACATATGCGGGGCTTTTCTGTGATGATAAC 3545
DB 1535 CTGGCTGCTGCCAACTCAGTGTTCGGCCGCTGGGATGAG- ---ACGAAGGGGAGGCAAC 1591
QY 3546 ATCCAGCTTCCCTCAACTCTACTTCTAGATTTGATTTAATTTACTTAAATGCTCGACAA 3605
DB 1592 ATTGACTTATGCCCCCACCCTTTGTGCGGCTTCGACATGATCTTCATTCGTAAGGATGAG 1651


```
QY 3606 CCAGACGAGCAAAACGATCGTGGTCTGCCAGGCGATCTCGTGGCTTTACAC 3656
Db 1652 CACAATGAGGAGAGGGATGTGATGCTGGCCAAAGCATGTTCATCACTCTGCAC 1702

RESULT 9
US-09-949-016-2621
; Sequence 2621, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2621
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2621

Query Match 4.1%; Score 180; DB 3; Length 3000;
Best Local Similarity 51.2%; Pred. No. 2.4e-47;
Matches 479; Conservative 0; Mismatches 445; Indels 12; Gaps 2;

QY 3020 GGGCATTTATGATAGACTTTCAAGTCTGCTCAAGCATTTGGGAGCTTGAAGATAT 3079
Db 1625 GGAATTTTACGAAAAGCTGGCAGCTTCAATCGCCCGAGAAATATACGGGCATGAAGATGT 1684

QY 3080 TAAAAGGCTCTTTTCCAGCTCTTTCGTGGGAGGCTAAGAAAATTCATCTGGAGC 3139
Db 1685 GAAGAAGGCACCTGCTCTCTCTAGTCGGGGGTGG---ACCAGTCTCTCGAGGCAT 1741

QY 3140 ATCTTTCCGAGGTGACATCAATGTTTTACTTTGGGGACCTCGGTACCAGTAATCTCA 3199
Db 1742 GAAATCCGGGGCAACATCAACATCTGTCTGATGGGGATCTGTGTGCCCAAGTCTCA 1801

QY 3200 GCTGCTTCATGATGTGCCAAGATAGCTCTCTGTGGAAATCTACACTAGTGGGGAGGAG 3259
Db 1802 GCTCTCTGTATACATTTGATCGACTGGCGCTCTGCAGCCAGTACACAAACGCGGGGCTC 1861

QY 3260 TTCCGGGTTGGCTCACAGCGTATGTAACGAAGGATCCAGAACTCGAGAGACGGTATT 3319
Db 1862 CTGAGAGTGGGGCTTACGGCAGCTGTCTGAGAGACTCGTGAGTGGAGAACTGACCTT 1921

QY 3320 GGAGAGCGGAGCTTTGGTCTTTAGTGATCGTGGGATATGCTGATCGATGAGTTCCGACAA 3379
Db 1922 AGAGGTGGGGCTCTGTGTGCTGCTGCCAGGCTGTGTCTGATGATGAGTTCCGACAA 1981

QY 3380 AATGTCGTGAATAATGCCGAGCATGCTTCAATGAGGTAATGGAGCAACAAACGGTATCTGT 3439
Db 1982 GATGGTGTAGGCCGACCGCACAGCCATCCACGAGGTCATGGAGCAGACCACTCTCCAT 2041

QY 3440 AGCCAAAGGGGTATCATTTGCTCGTGAACGCTCCGACGCTGCTCTTGCATGTGCMAA 3499
Db 2042 TGCCAAAGGCGGCAATTTCTACCAACACTCAATGCCCGCTGCTCTCCATCTGGCTGCCGCAA 2101

QY 3500 TCTAGTGGGTCCCGATCAATAGCGGGCTTTCTGTGATTTGATAAATCCAGCTTCTCTCC 3559
Db 2102 CCTTGCTACGGGGCTTACAACTCTCGCCGAGCTGGAGCAGAAACATACAGCTACCTGC 2161

QY 3560 AACTCTACTTTCTAGATTGAATTAATTTACTTAATGCTGACAAACAGCAGCAAAA 3619
Db 2162 TGCACTGCTCTCCGGTTTGACCTCTCTGGCTGATTCAGGACCGGCGCCGACGAGACAA 2221
```

```
QY 3620 CGATCGTCTCTGCCAGGCGATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAA 3679
Db 2222 TGACCTACGGTTGGCCAGCACATCACTATGTGCAACAGCAGCCGCGAGCCCCCTC 2281

QY 3680 GCAGGAGCGCTTAGATCTACAAACACTTACCGCTATATCACTATGCTCTCGTCAGCATGT 3739
Db 2282 CCASTTTGAACCTCTGGACATGAAGCTCATGAGGGTTACATAGCCATGTGCCGAGAA 2341

QY 3740 ACATCCTACATTAAGTGAAGCTGCTGAAGATTTGATTAATGGCTATGTGAGATCGG 3799
Db 2342 GCAGCCCATGTGCCAGAGTCTCTGGCTGACTACATCACAGCAGCATACGTGGAGATGAG 2401

QY 3800 CCAAAGGGCAACTTTCTTGGAGCAGTAAAGAGTAAACAGCCACACCTCGCAACT 3859
Db 2402 GCGAGAGG-----CTTGGGCTAGTAAGGATGCCACCTATATCTTCTGCCCGGACCT 2452

QY 3860 CGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTCTGAAGTGGTAGA 3919
Db 2453 GCTGGCTATCTCGCCCTTTCCACTGCTCTGGCAGCTCTGAGAAATGGTGGATGTGGTGA 2512

QY 3920 GAAAAGTTGATGCAGCAGAAAGCTGTGCGCTTTTGA 3955
Db 2513 GAAAGAGATGTGAATGAAGCCATCAGGCTAATGGA 2548
```

RESULT 10

```
US-09-248-796A-4570
; Sequence 4570, Application US/09248796A
; Patent No. 6747137
```

```
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
```

```
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
```

```
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
```

```
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
```

```
; SEQ ID NO 4570
; LENGTH: 996
```

```
; TYPE: DNA
```

```
; ORGANISM: Candida albicans
```

```
US-09-248-796A-4570
```

```
Query Match 4.0%; Score 173.2; DB 3; Length 996;
```

```
Best Local Similarity 56.5%; Pred. No. 1.9e-45;
```

```
Matches 322; Conservative 0; Mismatches 248; Indels 0; Gaps 0;
```

```
QY 2992 AAAAACTTAAAGAGCTGTCCAAGCTCCCGGCAATTTATCATAGACTTTCAAGGTCGCTGG 3051
Db 407 ACAAATCGCAGAGCTTTTACTTCAAGGTGATGTTTACAATAAGTTGGCCAAATCCATTG 466
```

```
QY 3052 CTCCTCAAGCAATTTGGGAGCTTGAAGATATTTAAAAAGGGTCTCTTTGCCAGCTCTTTGGTG 3111
Db 467 CTCGGGAATTTATGGTCACTTGGATGTTTAAAGAAAATTTTATTTATTGTTATTATGTTG 526
```

```
QY 3112 GGAAGGCTAAGAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTG 3171
Db 527 GTGTCAACCAAGAAAATCGGTGATGGTTTGAATAATAGAGGTGATATTAATGTTTGTCTTA 586
```

```
QY 3172 TTGGGGACCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTC 3231
Db 587 TGGGTGATCCGGGGGTGCTAAATCAAAATTTGAAAGCCATTTGGTAAGATTTGCTCCAA 646
```

```
QY 3232 GTGGAATCTACACTAGTGGGCGAGGAAGTTCCGCGGTTGGGCTGACAGCGTATGTAAACGA 3291
Db 647 GATCGGCTCTACTACTACTGCTAGAGGTTCTTCTGGTGTAGGGTTGACGGCAGCTGTTATGA 706
```

```
QY 3292 AGGATCCAGAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTTAGTGATCGTG 3351
```

Db 707 GAGATCCAAATTACTGATGAATGTTGGAGGTGGGCTTTTGGTGTGGCCGATACG 766
Qy 3352 GGATATGCTGATCGATGAGTTTCGACAAATGTTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db 767 GTATTGTGTTGATGATGATTTGATAAATGGACGAAAGCGACAGAACGGCGATTATG 826
Qy 3412 AGGTAATGGAGCACAAACGGTATCTGTAGCAAAAGGGGTATCATTCCTCGCTGACG 3471
Db 827 AAGTTATGGAAACAACAACCATCTCTATTGCAAGCTGGTATCACTACTACTTTGAATG 886
Qy 3472 CTCGGAGCTGTCTTGGCATGTCAAAATCTAGTGGTCCCGATACAAATGCGGCGCTTT 3531
Db 887 CCAGAACTTCGATCTTGGCAGCAGCCAAATCCATTTATGATGATGATATAATCAAGATTGT 946
Qy 3532 CTGTGATGTAACATCCAGCTTCTCCAA 3561
Db 947 CTCACATGAATAATCAATTTACCAGCCA 976

RESULT 11
US-09-964-899-48
; Sequence 48, Application US/09964899
; Patent No. 690367
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; FILE REFERENCE: Alzheimer's Disease Using Drosophila Melanogaster
; CURRENT APPLICATION NUMBER: US/09/964,899
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 3378
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-964-899-48

Query Match 3.7%; Score 160.8; DB 3; Length 3378;
Best Local Similarity 50.9%; Pred. No. 5.1e-41;
Matches 381; Conservative 0; Mismatches 367; Indels 0; Gaps 0;
Qy 2940 GCTGGGTATCATGAAAGTGATACCTTCAGAAAGCTGCTAATGAAGCAAGATTCAAAAACCTT 2999
Db 1333 GCCAAGAAGGACAAAGGTTGCTGTAGGGGAAGTACGAGTGAAGATGTGAAGATGATC 1392
Qy 3000 AAAGAGCTGTCGAAGTCCCGGGCATTTATGATAGACTTTCAAGGTGCTGGCTCCAAAGC 3059
Db 1393 ACTAGGCTCTCCAAAGGATCAGCAGATCGGAGAGAGATCTTTGCGCAGATTGCTCTCTCC 1452
Qy 3060 ATTTGGGAGCTTCAAGATATTAAGAGGCTTCTTTGCGCAGCTCTTTGGTGGAGGCT 3119
Db 1453 ATCTATGGTTCATGAAGACATCAAGAGAGGCTGGCTCTGCGCCCTGTTGCGAGGGAGGCC 1512
Qy 3120 AAGAAAATTCATCTGAGCATCTTTCCGAGGTGACATCAATGTTTACTTTGTTGGGAC 3179
Db 1513 AAAAACCAGGTGGCAGCAGAGGTACGTTGGTGATATCAAGCTGCTTTGTCGGGAGAC 1572
Qy 3180 CTGGGTACCAAGTAAATCTCAGCTGCTTTCAGTATGTGCAAGATAGCTCTCTCGTGAATC 3239
Db 1573 CTGGGCACAGCAAGTCGAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATC 1632
Qy 3240 TACACTAGTGGGAGGAAATTCGCGGTTGGGCTGACAGCGTATGTAAAGAGATCCA 3299
Db 1633 TTCAACACTGGCCAGGGGGCTGGGCTGTTGGGCTCAGCGGTATGTCCAGCGGACCCCT 1692
Qy 3300 GAAACTCGAGAGACGTTATGGAGAGCGGAGCTTTGGTCTTATGATCGTGGGATATGC 3359

Db 1693 GTACAGAGGAGTGGACCTTGGAGGCTGGGGCCCTGTGTTCTGGCTGACCCGAGGATGTGT 1752
Qy 3360 TGATTCGATGAGTTTCGACAAAATGTTCTGATAAATGCCCGAAGCATGCTTCATGAGGTAAATG 3419
Db 1753 CTCATTGATGAATTTGACAGATGAATGACCAAGACAGACCAAGCATCCATGAGGCCATG 1812
Qy 3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTTGCTCGCTGAAACGCTCGGACG 3479
Db 1813 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACTCCCTCGAGGCTCGCTGC 1872
Qy 3480 TCTGTCTTCGATCTGCAATCTAGTGGGTCCCGATACATGCGCGCTTCTGTGATTT 3539
Db 1873 ACGTCAATTTGTCGCGCAACCCCATAGGAGGGGCTACGACCCCTCGCTGACTTTCTCT 1932
Qy 3540 GATAACAATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTTACTTAATGCTC 3599
Db 1933 GAGAACGTGGACCTCAGAGGCCATCATCTCAGCTTTGACATCTGTGTGGTGAGG 1992
Qy 3600 GACAAACACAGCAGCAAAACGATCGTCTTCGCGAGGCATCTCGTGGCTTTTACACTAT 3659
Db 1993 GACACCGTGACCCAGTCCAGGACGAGATGCTGGCCGCTTCGTGGTGGGCGACCACTC 2052
Qy 3660 GAAAACTATGAAGTTTCAAGCAGGACG 3687
Db 2053 AGACACCACCCAGCAACAAGGAGGAGG 2080

RESULT 12
US-09-976-594-323
; Sequence 323, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 323
; LENGTH: 3445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1723834CB1
US-09-976-594-323

Query Match 3.7%; Score 160.8; DB 3; Length 3445;
Best Local Similarity 50.9%; Pred. No. 5.1e-41;
Matches 381; Conservative 0; Mismatches 367; Indels 0; Gaps 0;
Qy 2940 GCTGGGTATCATGAAAGTGATACCTTCAGAAAGCTGCTAATGAAGCAAGATTCAAAAACCTT 2999
Db 1387 GCCAAGAAGGACAAACAAAGGTTGCTGTAGGGGAAGTACCGCATGAAGATGTGAAGATGATC 1446
Qy 3000 AAAGAGCTGTCGAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTGCTGGCTCCAAAGC 3059
Db 1447 ACTAGGCTCTCCAAAGATCAGCAGATCGGAGAGAGATCTTTGCCAGCATTTGCTCTCTTC 1506
Qy 3060 ATTTGGGAGCTTCAAGATATTAAGAGGCTTCTTTTGGCAGCTCTTTGGTGGGAGGCT 3119
Db 1507 ATCTATGGTTCATGAGACATCAAGAGGCTGGCTCTGGCCCTGTTGCGAGGGAGGCC 1566
Qy 3120 AAGAAAATTCATCTGAGCATCTTTCCGAGGTGACATCAATGTTTACTTTTGTGGGAC 3179
Db 1567 AAAAACCAGGTGGCAAGCAGCAGGTACGTTGGTGATATCAACGCTGCTTTGTGCGGAGAC 1626
Qy 3180 CTGGGTACCAAGTAAATCTCAGCTGCTTCAGTATGTGCAAGATAGCTCTCTCGTGAATC 3239

THIS PAGE BLANK (USPTO)

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New). Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

THIS PAGE BLANK (USPTO)

2461 TGGTTATCCGGTCGAGTCTTATCATACCTGAAATTAAGGGGCGCTTCTTCAAATGTTTAG 2520
Db |||||
2461 TGGTTATCCGGTCGAGTCTTATCATACCTGAAATTAAGGGGCGCTTCTTCAAATGTTTAG 2520
Qy |||||
2521 TGTGTGGTCACTCGCTCGCTAGTATGACATGTTGTTAAAGGGCGGTTGAGGAGCCACAA 2580
Db |||||
2521 TGTGTGGTCACTCGCTCGCTAGTATGACATGTTGTTAAAGGGCGGTTGAGGAGCCACAA 2580
Qy |||||
2581 GGTGTGAAAGCCAGAAATGTGACAGACGGAATGCTATGCTCTTATTCACAAATCGATGCA 2640
Db |||||
2581 GGTGTGAAAGCCAGAAATGTGACAGACGGAATGCTATGCTCTTATTCACAAATCGATGCA 2640
Qy |||||
2641 CTTTGTCAAATAAGCAGATAGTGCCTTTTCAAGAACTCCAGATGCCATTCCTCGAAGGAG 2700
Db |||||
2641 CTTTGTCAAATAAGCAGATAGTGCCTTTTCAAGAACTCCAGATGCCATTCCTCGAAGGAG 2700
Qy |||||
2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG 2760
Db |||||
2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACTATGTTGATGCTGTGAAGCCTG 2760
Qy |||||
2761 GAGATCGTATGAGGTAAACAGAGTTCAGAGCCATGCGAGTTCGAGTTGGTCCGAATC 2820
Db |||||
2761 GAGATCGTATGAGGTAAACAGAGTTCAGAGCCATGCGAGTTCGAGTTGGTCCGAATC 2820
Qy |||||
2821 AACGAACATTACGAGCATTTGATATAGACCTACATCGATTCGCGTCAAGAGTCTG 2880
Db |||||
2821 AACGAACATTACGAGCATTTGATATAGACCTACATCGATTCGCGTCAAGAGTCTG 2880
Qy |||||
2881 ACAGGGGTGACCTGCAACTGAGATCTTATGAGATGATGATGAGATGATATGATG 2940
Db |||||
2881 ACAGGGGTGACCTGCAACTGAGATCTTATGAGATGATGATGAGATGATATGATG 2940
Qy |||||
2941 CTGGGTATCATGAAGTGAATCTTCAAGAGCTGCTAAATGAAGCAAGATTCAAAACCTTA 3000
Db |||||
2941 CTGGGTATCATGAAGTGAATCTTCAAGAGCTGCTAAATGAAGCAAGATTCAAAACCTTA 3000
Qy |||||
3001 AAGAGCTGTCCAAGTCCCGGCAATTTATGATAGACTTTCAAGGTGCTGGCTCCAAGCA 3060
Db |||||
3001 AAGAGCTGTCCAAGTCCCGGCAATTTATGATAGACTTTCAAGGTGCTGGCTCCAAGCA 3060
Qy |||||
3061 TTTGGAGCTTGAAGATATTAAGAGGCTCTTTGGCAGCTCTTTGGTGGGAGGCTTA 3120
Db |||||
3061 TTTGGAGCTTGAAGATATTAAGAGGCTCTTTGGCAGCTCTTTGGTGGGAGGCTTA 3120
Qy |||||
3121 AGAAATTCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTACTTTGGGGACC 3180
Db |||||
3121 AGAAATTCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTACTTTGGGGACC 3180
Qy |||||
3181 CTGGTACCAAGTAACTCAGCTGCTTCAAGTATGTCACAGATAGCTCCTCGTGAATCT 3240
Db |||||
3181 CTGGTACCAAGTAACTCAGCTGCTTCAAGTATGTCACAGATAGCTCCTCGTGAATCT 3240
Qy |||||
3241 ACACTAGTGGGCGAGGAGTTCGGCGGTTGGGCTGACAGCTATGTAAAGAGATCCAG 3300
Db |||||
3241 ACACTAGTGGGCGAGGAGTTCGGCGGTTGGGCTGACAGCTATGTAAAGAGATCCAG 3300
Qy |||||
3301 AAACCTGAGAGCGGTATTTGAGAGCGGAGCTTTGGTCTTAGTATGCTGGGATATGCT 3360
Db |||||
3301 AAACCTGAGAGCGGTATTTGAGAGCGGAGCTTTGGTCTTAGTATGCTGGGATATGCT 3360
Qy |||||
3361 GTATCGATGAGTTCGACAAATGCTGATAATGCCGGAAGCATGCTTCATGAGTAAATCG 3420
Db |||||
3361 GTATCGATGAGTTCGACAAATGCTGATAATGCCGGAAGCATGCTTCATGAGTAAATCG 3420
Qy |||||
3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGTATCAATGCTCGCTGAACGCTCGGACGT 3480
Db |||||
3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGTATCAATGCTCGCTGAACGCTCGGACGT 3480
Qy |||||
3481 CTGTCTTGTGATGTGCAAAATCTTAGTGGGTCCGATACAAATGCGCGCCCTTCTGTGATG 3540
Db |||||
3481 CTGTCTTGTGATGTGCAAAATCTTAGTGGGTCCGATACAAATGCGCGCCCTTCTGTGATG 3540

3541 ATAAATCOAGCTTCTCCTCAACTCTACTTCTTAGATTTGATTAATTAATTAATTAATTAATTA 3600
Db |||||
3541 ATAAATCOAGCTTCTCCTCAACTCTACTTCTTAGATTTGATTAATTAATTAATTAATTAATTA 3600
Qy |||||
3601 ACAACACAGAGCAGCAAAACGATCCTGCTCGCCAGGACATCTCGTGGCTTTACACTATG 3660
Db |||||
3601 ACAACACAGAGCAGCAAAACGATCCTGCTCGCCAGGCACTCTCGTGGCTTTACACTATG 3660
Qy |||||
3661 AAAAATATGAAGTTTCAAAGCAGGACGCTTTAGATCTACAAACACTTACCCTGATATCA 3720
Db |||||
3661 AAAAATATGAAGTTTCAAAGCAGGACGCTTTAGATCTACAAACACTTACCCTGATATCA 3720
Qy |||||
3721 CCTATGCTCGTCAGCATGTACATCTACATTAAGTGAAGAGTCTGTAAGATTTGATTA 3780
Db |||||
3721 CCTATGCTCGTCAGCATGTACATCTACATTAAGTGAAGAGTCTGTAAGATTTGATTA 3780
Qy |||||
3781 ATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTTGGAGAGCAGTAAAGAGGTGATAA 3840
Db |||||
3781 ATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTTGGAGAGCAGTAAAGAGGTGATAA 3840
Qy |||||
3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
Db |||||
3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
Qy |||||
3901 GATTTTCTGAAGTGTAGAGAAAGTTGATGACAGAGAGCTGTGCGCTTTTAGAGCTCG 3960
Db |||||
3901 GATTTTCTGAAGTGTAGAGAAAGTTGATGACAGAGAGCTGTGCGCTTTTAGAGCTCG 3960
Qy |||||
3961 CTTTTCAGCMACTCTACTGATCATGCAACAGGAGTACGATGATGATGATGATGATGATGATG 4020
Db |||||
3961 CTTTTCAGCMACTCTACTGATCATGCAACAGGAGTACGATGATGATGATGATGATGATGATG 4020
Qy |||||
4021 CTGGAGTGTGCGGCAAGCAGGCTATTCGTGCGGCGCAACTTGTAGTGTCTCTCGGAGAGC 4080
Db |||||
4021 CTGGAGTGTGCGGCAAGCAGGCTATTCGTGCGGCGCAACTTGTAGTGTCTCTCGGAGAGC 4080
Qy |||||
4081 TTATAGCAGATAAAATTTCACTGCGAGCTCTCTGGCTTGAAGACCACTAGCTAGCTTTCTTG 4140
Db |||||
4081 TTATAGCAGATAAAATTTCACTGCGAGCTCTCTGGCTTGAAGACCACTAGCTAGCTTTCTTG 4140
Qy |||||
4141 AGGATATCGGAGCCAAAGCAGTGTGGAGTGTAGTTTGCAGGATATTAAGATGCTCTCG 4200
Db |||||
4141 AGGATATCGGAGCCAAAGCAGTGTGGAGTGTAGTTTGCAGGATATTAAGATGCTCTCG 4200
Qy |||||
4201 GTAGGCTTCAAGGAGAGCTTTCTTACTGTCCATGTCACATAGTCAAGAGAGTTTGAG 4260
Db |||||
4201 GTAGGCTTCAAGGAGAGCTTTCTTACTGTCCATGTCACATAGTCAAGAGAGTTTGAG 4260
Qy |||||
4261 ACAGTTTCTAACTGTTTCAATCCATGAGCTATAACTCTGAACGAAAGGGAACCTCCAG 4320
Db |||||
4261 ACAGTTTCTAACTGTTTCAATCCATGAGCTATAACTCTGAACGAAAGGGAACCTCCAG 4320
Qy |||||
4321 TTTCCATGCGCAATTTCCAGAGCTCGC 4348
Db |||||
4321 TTTCCATGCGCAATTTCCAGAGCTCGC 4348

RESULT 2

US-10-768-511-5
; Sequence 5, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/10/768,511
; CURRENT FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: 60/196,001

; PRIOR FILING DATE: 2000-04-07									
; NUMBER OF SEQ ID NOS: 34									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 5									
; LENGTH: 4348									
; TYPE: DNA									
; ORGANISM: Physcomitrella patens									
US-10-768-511-5									
Query Match 100.0%; Score 4348; DB 7; Length 4348;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 4348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	ATGGCGCGCGCAGCTCAGCTGAGGAATTCACCTCTCTTGTCTCGACGGTCCATCTTT	60						
Db	1	ATGGCGCGCGCAGCTCAGCTGAGGAATTCACCTCTCTTGTCTCGACGGTCCATCTTT	60						
Qy	61	TTTGGTTTTAGTTTGAAATCTTGATCGTGGAGTTGAGAAAGGGCGGTTCTGTGTCT	120						
Db	61	TTTGGTTTTAGTTTGAAATCTTGATCGTGGAGTTGAGAAAGGGCGGTTCTGTGTCT	120						
Qy	121	TGAGGTGTTCTGTTGATTTGTCATGGAATAATGATGCATCTGACATTTGAGCGG	180						
Db	121	TGAGGTGTTCTGTTGATTTGTCATGGAATAATGATGCATCTGACATTTGAGCGG	180						
Qy	181	TGTCGTCCCATATCTTCGCAATCTGAAGAGTGTCTACGCCATTGGCGCAAGTAACAT	240						
Db	181	TGTCGTCCCATATCTTCGCAATCTGAAGAGTGTCTACGCCATTGGCGCAAGTAACAT	240						
Qy	241	CACCGAGCTTCGAATGAGCCTCACTCGTGGCGGGCGGAGGCGCTGACGGAGAGCC	300						
Db	241	CACCGAGCTTCGAATGAGCCTCACTCGTGGCGGGCGGAGGCGCTGACGGAGAGCC	300						
Qy	301	CTACATCTCAGTTTGGAGGAGGAGAAACGGATTCGCTCGTAGAGAGGATC	360						
Db	301	CTACATCTCAGTTTGGAGGAGGAGAAACGGATTCGCTCGTAGAGAGGATC	360						
Qy	361	GATCTCGCAGTTTGGCAATCTGTTATAGTTTCCCTTACGATCGGGGACTCCTGGAA	420						
Db	361	GATCTCGCAGTTTGGCAATCTGTTATAGTTTCCCTTACGATCGGGGACTCCTGGAA	420						
Qy	421	CTCTCGGAATCTCAGTGGCTACTCCGGTTTACGTAACCCAGTCCGTATGGGTA	480						
Db	421	CTCTCGGAATCTCAGTGGCTACTCCGGTTTACGTAACCCAGTCCGTATGGGTA	480						
Qy	481	CCCCATCGTTCATCTGTGGCAGCCACAGTACAAACAGCGCAGTGGTTCACAG	540						
Db	481	CCCCATCGTTCATCTGTGGCAGCCACAGTACAAACAGCGCAGTGGTTCACAG	540						
Qy	541	GGAGGCTCTACATCGGAGAGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCTT	600						
Db	541	GGAGGCTCTACATCGGAGAGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCTT	600						
Qy	601	CAAGGAACTAGTGTGATGGCGTCCCTCTGAATCTGTGAGCCAGATGACATTTGG	660						
Db	601	CAAGGAACTAGTGTGATGGCGTCCCTCTGAATCTGTGAGCCAGATGACATTTGG	660						
Qy	661	GTGGAGATATGCTTATGTTTGGGGAGCAATCTTAACATCCAGATGTCTTAGGGCGA	720						
Db	661	GTGGAGATATGCTTATGTTTGGGGAGCAATCTTAACATCCAGATGTCTTAGGGCGA	720						
Qy	721	TTCTGTCGATTTCTCCAAATTAATCTGTTCCGAGTGTCTCATGATCTTAAATCCAGTACATCC	780						
Db	721	TTCTGTCGATTTCTCCAAATTAATCTGTTCCGAGTGTCTCATGATCTTAAATCCAGTACATCC	780						
Qy	781	AGATCATAGAGGAGACTGTGAGCGGTGAGGAGATCTTAATATCGACATGTGAGACA	840						
Db	781	AGATCATAGAGGAGACTGTGAGCGGTGAGGAGATCTCTAAATATCGACATGTGAGACA	840						
Qy	841	TTTTATGACATGATCTGATCTATAGCAAAATTTGTTCCATACCCACTCGACATCTCC	900						
Db	841	TTTTATGACATGATCTGATCTATAGCAAAATTTGTTCCATACCCACTCGACATCTCC	900						

Qy	901	CCCTGTTGGACACTGAGTGTGAGAGTTGCTACCTTCTTACTACCAACGTTTGAGAAGC	960
Db	901	CCCTGTTGGACACTGAGTGTGAGAGTTGCTACCTTCTTACTACCAACGTTTGAGAAGC	960
Qy	961	ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACTGCGTGAACCTCAACCTT	1020
Db	961	ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACTGCGTGAACCTCAACCTT	1020
Qy	1021	CAGATATAGACAAATTTGTTTAAAGAAATGTTTCCCGTGCAGTCTCTATCATAC	1080
Db	1021	CAGATATAGACAAATTTGTTTAAAGAAATGTTTATCCGGTGCAGTCTCTATCATAC	1080
Qy	1081	CTGAAATTAAGGGCGCTCTTCAAATGTTTAGTGTGCTGCTCGCTCCGCTAGTTA	1140
Db	1081	CTGAAATTAAGGGCGCTCTTCAAATGTTTAGTGTGCTGCTCGCTCCGCTAGTTA	1140
Qy	1141	CAGTGTGTTAAAGGGCGGTTGAGGAGCCAAAGGTTGTAAGCCAGAAATGTGCAGCAC	1200
Db	1141	CAGTGTGTTAAAGGGCGGTTGAGGAGCCAAAGGTTGTAAGCCAGAAATGTGCAGCAC	1200
Qy	1201	GGAAATGCTATGCTCTTATTTCACAATCGATGCACCTTTTGCAAAATAGCAGATAGTGC	1260
Db	1201	GGAAATGCTATGCTCTTATTTCACAATCGATGCACCTTTTGCAAAATAGCAGATAGTGC	1260
Qy	1261	TTCAAGAACTCCAGATGCTGCTGAAAGCTGAGAGATCGTATTTGAGGTAACAGGAGTTT	1320
Db	1261	TTCAAGAACTCCAGATGCTGCTGAAAGCTGAGAGATCGTATTTGAGGTAACAGGAGTTT	1320
Qy	1321	TATCAACACATATGTTGATGCTGTAAGCTGAGAGATCGTATTTGAGGTAACAGGAGTTT	1380
Db	1321	TATCAACACATATGTTGATGCTGTAAGCTGAGAGATCGTATTTGAGGTAACAGGAGTTT	1380
Qy	1381	TCAAGGCCATCGCAGTTGAGTTCATGCGCGCGCAGCTCAGTGAAGAAATTCGACC	1440
Db	1381	TCAAGGCCATCGCAGTTGAGTTCATGCGCGCGCAGCTCAGTGAAGAAATTCGACC	1440
Qy	1441	TCCTTGTCTCGGAGGTTCCATCTTTTGGTTTTAGTTTGAATACTTGATCGTGA	1500
Db	1441	TCCTTGTCTCGGAGGTTCCATCTTTTGGTTTTAGTTTGAATACTTGATCGTGA	1500
Qy	1501	GTTGAGAAAAAGGGCGGTTTCGTTGATGAGTGTCTTGTGATGTTGTCATGGA	1560
Db	1501	GTTGAGAAAAAGGGCGGTTTCGTTGATGAGTGTCTTGTGATGTTGTCATGGA	1560
Qy	1561	ATAATGATGCACTTGACATTTGAGCGCGTGTCTGTCCTCATATCTTTCGCAATCTGAAGG	1620
Db	1561	ATAATGATGCACTTGACATTTGAGCGCGTGTCTGTCCTCATATCTTTCGCAATCTGAAGG	1620
Qy	1621	TGTCTACGCCATTGCGCAAGTAAACATCACCGAGCTTCGACAAATCGAGCTCACCGTGG	1680
Db	1621	TGTCTACGCCATTGCGCAAGTAAACATCACCGAGCTTCGACAAATCGAGCTCACCGTGG	1680
Qy	1681	CCGGCGGAGGCGCGTACGGCAGACCCCTTACATCTGCAAGTTCGAAAGGAGGAGAGAAA	1740
Db	1681	CCGGCGGAGGCGCGTACGGCAGACCCCTTACATCTGCAAGTTCGAAAGGAGGAGAGAAA	1740
Qy	1741	CGGATTCGCTCGTCTAGGAGGAGTTCGATCTCGAGTTTGAAGCAATTCGTTTATAGTT	1800
Db	1741	CGGATTCGCTCGTCTAGGAGGAGTTCGATCTCGAGTTTGAAGCAATTCGTTTATAGTT	1800
Qy	1801	CCCTTACGATCGGGGACTCTTGAACTCTCTGGAACCTCCAGTGGCTACTCCGGTTTACG	1860
Db	1801	CCCTTACGATCGGGGACTCTTGAACTCTCTGGAACCTCCAGTGGCTACTCCGGTTTACG	1860
Qy	1861	CTACCCAGTTCGCTATGGGTACCCCATCGTTTCCATCGTGGCAGCCACAGTACA	1920
Db	1861	CTACCCAGTTCGCTATGGGTACCCCATCGTTTCCATCGTGGCAGCCACAGTACA	1920
Qy	1921	AACAGCGCAGTGTGCTTCCCGAGGAGCGCTCTATCGGAGAGCGTCTCAAT	1980
Db	1921	AACAGCGCAGTGTGCTTCCCGAGGAGCGCTCTATCGGAGAGCGTCTCAAT	1980
Qy	1981	CCAGAGAACCCGGGCATCGATCTCTCTTCAAGGGAACCTTAGTGTGCGCTCCCTCTG	2040

1981 DB ||||| CCAGAGAACCGGCGATCGATCTCTTCAAGGGAACCTAGTGTGATGGCGCTCCCTCTG 2040
2041 QY ||||| AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAAATATGCTTATGTTTGGGGGACGAATG 2100
2041 DB ||||| AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAAATATGCTTATGTTTGGGGGACGAATG 2100
2101 QY ||||| TTAACATTCAGATGCTTTAGGGCGAATCGTTCGATTTCTCCAAATATATCGTTCGAGTG 2160
2101 DB ||||| TTAACATTCAGATGCTTTAGGGCGAATCGTTCGATTTCTCCAAATATATCGTTCGAGTG 2160
2161 QY ||||| CTCATGATCTTAATCCAGTACATCCAGATCATAGAGGAGACTGTGGAGCGGTGAGGAG 2220
2161 DB ||||| CTCATGATCTTAATCCAGTACATCCAGATCATAGAGGAGACTGTGGAGCGGTGAGGAG 2220
2221 QY ||||| ATACTCTAAATATCGCATGTGAGACATTTATGACCATGATCTGATCTATATCGCAAAA 2280
2221 DB ||||| ATACTCTAAATATCGCATGTGAGACATTTATGACCATGATCTGATCTATATCGCAAAA 2280
2281 QY ||||| TTGTTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTTCAGGAAGTTGCTA 2340
2281 DB ||||| TTGTTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTTCAGGAAGTTGCTA 2340
2341 QY ||||| CCTCTTTATACCAAGTTTGAGAGCATATTCAGGCCAGACCTTTCAATCTCAAAGCAT 2400
2341 DB ||||| CCTCTTTATACCAAGTTTGAGAGCATATTCAGGCCAGACCTTTCAATCTCAAAGCAT 2400
2401 QY ||||| CGGTGCACATGCGTGAACCTCAACCTTCAGATATAGACAAATTTGGTTCTGTTAAAGGAA 2460
2401 DB ||||| CGGTGCACATGCGTGAACCTCAACCTTCAGATATAGACAAATTTGGTTCTGTTAAAGGAA 2460
2461 QY ||||| TGTTTATCCGGTGCAGTCTTATCATACCTGAAATTAAGGGGGCTTCTTCAAATGTTTAG 2520
2461 DB ||||| TGTTTATCCGGTGCAGTCTTATCATACCTGAAATTAAGGGGGCTTCTTCAAATGTTTAG 2520
2521 QY ||||| TGTGTGCTACTGCTCCCTAGTTTACAGTTGTTTAAAGGGCGGTTGAGGAGCCCAACAA 2580
2521 DB ||||| TGTGTGCTACTGCTCCCTAGTTTACAGTTGTTTAAAGGGCGGTTGAGGAGCCCAACAA 2580
2581 QY ||||| GGTGTGAAAGCCAGAAATGTCAGCAGGAATGCTATGCTCTTATTCACAAATCGATGCA 2640
2581 DB ||||| GGTGTGAAAGCCAGAAATGTCAGCAGGAATGCTATGCTCTTATTCACAAATCGATGCA 2640
2641 QY ||||| CTTTTCGAAATAAGCAGATAGTGGTCTTCAAGAACTCCAGATGCCATTCCTGGAAGAG 2700
2641 DB ||||| CTTTTCGAAATAAGCAGATAGTGGTCTTCAAGAACTCCAGATGCCATTCCTGGAAGAG 2700
2701 QY ||||| AGACTCCACACACAGTCAGATGTTTATACAACTATGTTGATGCTGTGAAAGCCTG 2760
2701 DB ||||| AGACTCCACACACAGTCAGATGTTTATACAACTATGTTGATGCTGTGAAAGCCTG 2760
2761 QY ||||| GAGATCGTATTGAGTTAAGAGTTTCAAGGCCATGCGAGTTGCTCCGAATC 2820
2761 DB ||||| GAGATCGTATTGAGTTAAGAGTTTCAAGGCCATGCGAGTTGCTCCGAATC 2820
2821 QY ||||| AACGAACTTACAGACATGTTATAGACCTACATCGATTGCGTCAGCTCAAGAAAGTCTG 2880
2821 DB ||||| AACGAACTTACAGACATGTTATAGACCTACATCGATTGCGTCAGCTCAAGAAAGTCTG 2880
2881 QY ||||| ACAGGGGTGCACTGCAAACTGAAAGTCTTATGAGATGGAATAAGGAATGATATGATG 2940
2881 DB ||||| ACAGGGGTGCACTGCAAACTGAAAGTCTTATGAGATGGAATAAGGAATGATATGATG 2940
2941 QY ||||| CTGGGTATCATGAAAGTGATCTTCAGAACTGCTTAATGAAGCAAGATTCAAAACCTTA 3000
2941 DB ||||| CTGGGTATCATGAAAGTGATCTTCAGAACTGCTTAATGAAGCAAGATTCAAAACCTTA 3000
3001 QY ||||| AAGAGCTGCTCAAGCTCCCGGCAATTTATGATAGACTTTTCAAGTCTGCTCCGAAGCA 3060
3001 DB ||||| AAGAGCTGCTCAAGCTCCCGGCAATTTATGATAGACTTTTCAAGTCTGCTCCGAAGCA 3060
3061 QY ||||| TTTGGGAGCTTGAAGATATTTAAAGGGTCTTCTTTTCCAGCTCTTTTGGTGGGAAGGCTA 3120

3061 DB ||||| TTTGGGAGCTTGAAGATATTTAAAGGGTCTTCTTTTCCAGCTCTTTTGGTGGGAAGGCTA 3120
3121 QY ||||| AGAAAAATTCATCTGAGGACATCTTTCCGAGGTGACATCAATGTTTATCTTTGTTGGGAGC 3180
3121 DB ||||| AGAAAAATTCATCTGAGGACATCTTTCCGAGGTGACATCAATGTTTATCTTTGTTGGGAGC 3180
3181 QY ||||| CTGGTACAGTAAATCTAGCTGCTTCAATATGTCACAGATAGCTCTCTGTTGAATCT 3240
3181 DB ||||| CTGGTACAGTAAATCTAGCTGCTTCAATATGTCACAGATAGCTCTCTGTTGAATCT 3240
3241 QY ||||| ACACTAGTGGGAGGAAGTTCGGCGTTGGGCTGACAGCTATGTAAAGGAATCCAG 3300
3241 DB ||||| ACACTAGTGGGAGGAAGTTCGGCGTTGGGCTGACAGCTATGTAAAGGAATCCAG 3300
3301 QY ||||| AAACTCGAGAGCGGTATTTGGAGAGCGGAGCTTTGGTTCTTATGATGATCGTGGGATGCT 3360
3301 DB ||||| AAACTCGAGAGCGGTATTTGGAGAGCGGAGCTTTGGTTCTTATGATGATCGTGGGATGCT 3360
3361 QY ||||| GTATCGATGAGTTCGAAAAATGTCGTGATAATGCCCGAAGCATGCTTCATGAGGTAAATG 3420
3361 DB ||||| GTATCGATGAGTTCGAAAAATGTCGTGATAATGCCCGAAGCATGCTTCATGAGGTAAATG 3420
3421 QY ||||| AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTCGCTCGCTGAACGCTCGGACGT 3480
3421 DB ||||| AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTCGCTCGCTGAACGCTCGGACGT 3480
3481 QY ||||| CTGTCTTCATGTGCAAAATCCTAGTGGGTCCCGATACAAATGCGCGCTTCTTCTGATG 3540
3481 DB ||||| CTGTCTTCATGTGCAAAATCCTAGTGGGTCCCGATACAAATGCGCGCTTCTTCTGATG 3540
3541 QY ||||| ATAACTCAGCTTCTTCAACTCTTACTTCTAGATTTGATTTAAATTTACTTAATGCTCG 3600
3541 DB ||||| ATAACTCAGCTTCTTCAACTCTTACTTCTAGATTTGATTTAAATTTACTTAATGCTCG 3600
3601 QY ||||| ACAAAACACAGCAGCAAAACGATCGTCTGCTGCCAGGCAATCTCGTGGCTTTTACACTATG 3660
3601 DB ||||| ACAAAACACAGCAGCAAAACGATCGTCTGCTGCCAGGCAATCTCGTGGCTTTTACACTATG 3660
3661 QY ||||| AAAAATATGAAGTTTCAAAAGCAGGAGCGCTTAGATCTACAAACACTTACCGCGTATATCA 3720
3661 DB ||||| AAAAATATGAAGTTTCAAAAGCAGGAGCGCTTAGATCTACAAACACTTACCGCGTATATCA 3720
3721 QY ||||| CTTATGCTCGTCAGCATGTACCTTCAATTAAGTATGAAGTGTCTGAAGATTTGATTA 3780
3721 DB ||||| CTTATGCTCGTCAGCATGTACCTTCAATTAAGTATGAAGTGTCTGAAGATTTGATTA 3780
3781 QY ||||| ATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTGGAAGCAGTAAAGAGTGTATAA 3840
3781 DB ||||| ATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTGGAAGCAGTAAAGAGTGTATAA 3840
3841 QY ||||| CAGCCACACTCGGCAACTCGAAAGTATGATTCGATCAAGTGAAGCCCTAGCTCGAATGA 3900
3841 DB ||||| CAGCCACACTCGGCAACTCGAAAGTATGATTCGATCAAGTGAAGCCCTAGCTCGAATGA 3900
3901 QY ||||| GATTTTCTGAAGTGTGAGAAAGTTGATGTCAGCAGAAAGCTGTGCGCTTTTGAAGCTCG 3960
3901 DB ||||| GATTTTCTGAAGTGTGAGAAAGTTGATGTCAGCAGAAAGCTGTGCGCTTTTGAAGCTCG 3960
3961 QY ||||| CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGCATGGATCTTATCACGA 4020
3961 DB ||||| CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGCATGGATCTTATCACGA 4020
4021 QY ||||| CTGGAGTGTGGCCAGCGAGCGTATTTCTGTCGGGCCAACTTGTAGTGTCTCTCGAGAGC 4080
4021 DB ||||| CTGGAGTGTGGCCAGCGAGCGTATTTCTGTCGGGCCAACTTGTAGTGTCTCTCGAGAGC 4080
4081 QY ||||| TTATAGCAGATAAATTTCACTGCGAGCTCTCTCTGGCTTGAAGACAGGTCAAGTCTTCT 4140
4081 DB ||||| TTATAGCAGATAAATTTCACTGCGAGCTCTCTCTGGCTTGAAGACAGGTCAAGTCTTCT 4140
4141 QY ||||| AGGATATCCGGAGCCAAAGCAGGTGTGAAGTGTGAGTGTGAGGATATTTAAATTTGCTCTG 4200
4141 DB ||||| AGGATATCCGGAGCCAAAGCAGGTGTGAAGTGTGAGTGTGAGGATATTTAAATTTGCTCTG 4200

```
QY 4201 GTAGCTCCAGGAGAGGCTTTCTTACTGTCATGGTGACATAGTCAAGAGAGTTTGAG 4260
Db 4201 GTAGCTCCAGGAGAGGCTTTCTTACTGTCATGGTGACATAGTCAAGAGAGTTTGAG 4260
QY 4261 ACAGTTTCTTAAGTGTTCGAATCCATGAGCTATAAATCTTGACGAAGGGAACCTCCAG 4320
Db 4261 ACAGTTTCTTAAGTGTTCGAATCCATGAGCTATAAATCTTGACGAAGGGAACCTCCAG 4320
QY 4321 TTTCCTCATGCGCAATTCCTCAGAGCTCGC 4348
Db 4321 TTTCCTCATGCGCAATTCCTCAGAGCTCGC 4348

RESULT 3
US-10-425-114-28735
; Sequence 28735, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yinhua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaka, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 28735
; LENGTH: 2841
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: L1B4763-010-C6_FLI
US-10-425-114-28735
; Query Match 20.2%; Score 880.2; DB 7; Length 2841;
; Best Local Similarity 65.3%; Pred. No. 2.7e-265;
; Matches 1334; Conservative 0; Mismatches 688; Indels 21; Gaps 2;

QY 2217 GAGGATCTCTAATATCGACATGTGACGACATTTATGACCATGATCCTGATCTATACGCA 2276
Db 625 GGGGAGTCGCTCGACGTGATGCGCAGGCTGTTTGACCAACGACCGACCTCTACAGC 684
QY 2277 AAAATTGTTGATACCCACTCGACATCATCCCTGTTGGACACTGAGTGTGAGGAAGTT 2336
Db 685 AAGATGGTTTCGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGTCTCATGGACCTC 744
QY 2337 GCTACCTCTTTACTACCAAGTTTGGAGAGCATATTGAGGCCAGACCTTTCAATCTCAAA 2396
Db 745 GTCGCGGCATCGAGCGCTCTTCGAGAGACATTCGAGACAGGATCTACACCTCAAG 804
QY 2397 GCATCGGTGCACATGCGTGAACCTCAACCTTCAGATATAGACAAATTTGGTTTCTGTTAA 2456
Db 805 TCGTCCATTTGCTTGAGGATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAG 864
QY 2457 GGAATGGTTATCCGGTGCAGTTCTATCATACCTGGAATTAAGGGGGCTTCTTCAATGT 2516
Db 865 GGTATGATAATTAGATGCAGTCGGTCTATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGC 924
QY 2517 TTAGTGTGCTGCTCACTCGCTCGCTAGTTTACAGTTGTTTAAAGGGGGGTTGAGGAGCCA 2576
Db 925 CTGGTTTGTGGTTTCTACTCAGAGCCCGTCTATGGTTGATAGAGAGAGTAATCTGACCA 984
QY 2577 ACAAGGTGTGAAGAAGCCAGAAATGTCAGCAGCGGAATCTATGTCTCTTATTCACAAATCGA 2636
Db 985 CACATTTGTGAGAAAGAACATGTAAAGCCACAAATTTCTATGACCTTAGTGCACACAGA 1044
QY 2637 TGCATTTTGCAAATAAGACGAGATGTGCGTCTTCAAGAACTCCAGATGCCATTCCTGAA 2696
Db 2637 TGCATTTTGCAAATAAGACGAGATGTATTTAGCCACACAGTTATCTGATGAGCTGCAGAGAGTTA 2166

1045 TGCAGATTTTTCAGACAAGCAGATCATAAAGTTGTCAGGAAACACCCAGACGAGATACCAGAA 1104
2697 GGAGAGACTCCACACAGCTCAGCATGTGTTTATACACACTATGTTGTTGATGCTGTGAG 2756
1105 GGTGGCATCTCCACATACAGTTTAGTGTCTTGATGATGATTAAGCTTGTGTTGATGCTGGAAG 1164
2757 CCTCGAGATCGTATTTAGGTTAAACAGGAGTTTTCAGGCGCATGGCAGTTTCGAGTTGGTCCG 2816
1165 CCTCGAGATAGGGTTGAGATAACTTGGAAATATACAGAGCTAGATTTTGGATTTGGACCA 1224
2817 AATCAAGCAACATTTACGAGCAATTTGATAAGACCTACATCGATTCGATGCGTCAAGAG 2876
1225 ACTCAAGGACAGTGAAGTCTATATTCAGAGACATATATTGATTGCTTTCACATAAAGAG 1284
2877 TCTGACAGGGGTGACGTGCAAACTGAAGATCCTATGGAGATGGATGAAGAGATGATATG 2936
1285 ACAGACAAGTCTAGGCTTTCATGTGAGGACACCATGGATATTGATAATTTCTAAACCTAGC 1344
2937 TATGCTGGGTATCATGAAGTGATCTTCAGAAGCTGCTAATGAAGCAAGATTCAAAAA 2996
1345 AAATC-----TACTGAAGAGGATTTTCTTAGTGATGAAGTTGAGGAA 1386
2997 CTAAAGAGCTGTCCAGCTCCCGGCAATTTATGATAGACTTTCAAGGTCGCTGCTCCA 3056
1387 CTAAAGAGCTTTCGAAGTTCCTGATATCTATGAAGATTTGACTAGATCATTAGCTCCA 1446
3057 AGCATTTGGAGCTTTGAAGATATTAAAGGGTCTTCTTTGCGAGCTCTTTGGTGGGAG 3116
1447 AACATATGGAGTTGGATGATGTCAAAAGAGGTCTCTTTGCGAGCTTTTTCGGCGGTAA 1506
3117 GCTAAGAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTCTTACTGTTGG 3176
1507 CCCTTGAAGCTTCTCTTCTGGAGCTAGTTTTCGGGGTGACATCAATATTTTACTTGTGGG 1566
3177 GACCTCGTACCAAGTAAATCTCAGCTGCTTCAGTATGTGCAAGATAGCTCTCTCGTGA 3236
1567 GACCTCGAACAAGTAAATCCAGCTTCTCAGTACATGCAATAAATGCTCTCTCGTGT 1626
3237 ATCTACACTAGTGGCGGAGGAAGTTTCGGCGTTGGGCTGACAGCGTATGTAAAGAGAT 3296
1627 ATCTATACGAGTGGTAGAGGAAGTTCTGCTGTGTTGCTTACTGCTTATGTACCAAGAC 1686
3297 CAGAAACTCGAGACCGGTATTTGAGAGCGAGCTTTGGTTCTTAGTATCGTGGGATA 3356
1687 CCGTAGCTGGCGAAACTGTTCTAGAAAGTGAGACCTGTTTGTAGTGACAAAGGTGT 1746
3357 TCGTGTATCGATGAGTTTCGACAAATGCTGATAATGCCGGAAGCATGCTTCATGAGGTA 3416
1747 TGTGCTAGATAGTGTGATGAAGATGCTGATAATGCCGGAAGCATGTTACACGAGGTG 1806
3417 ATGGAGCAACAAACCGTATCTGTAGCAAAAGGGGTATCATTTGCTCGCTGAACGCTCGG 3476
1807 ATGGAACAGCAGACAGTATCCATTCGAAGGCTGGAATAATTGCAATCTTTAAACGCTAG 1866
3477 AGTCTGCTCTGCTGATGTCGAATCCTAGTGGTCCGATCAATGCGCGGCTTTCTGTCG 3536
1867 ACATCTGCTCGCATGTGCCAATCTCTGAAATCACGTTTACATCAAGGCTCTCTGTA 1926
3537 ATTGATAACATCCAGCTTCTCCAACTCTACTTCTCTAGATTTTGAATTTAATTTACTTAATG 3596
1927 ATTGACAACATCCACTTAGCGCCACGCTACTTTCAAGATTCGACCTGATTTATCTATC 1986
3597 CTCGACAAACCGAGCAGCAAAACGATCGTCTGTCGCGAGGATCTCGTGGCTTTACAC 3656
1987 TTGGACAAGCGCGATGAGCAAACTGATAGCGCTCGCAAGCATATTTGTTCTGTTGCAT 2046
3657 TATGAAACATTAAGAGTTTCAAGCAGGAGCGCTTAGATCTACAAACACTTTACCGGTAT 3716
2047 TTTGAGAAATCCAAATTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCTAC 2106
3717 ATCACCTATGCTCGTCAGCATGTACATCTCTACATTAAGTGAATGAAGCTGCTGAAGATTG 3776
2107 ATAACTATGCAAGGAAGTATATTAGCCACACAGTTATCTGATGAGCTGCAGAGAGTTA 2166
```


APPLICANT: BOHNERT, HANS J.
APPLICANT: VAN THIELEN, NOCHA
APPLICANT: CHEN, RUYING
APPLICANT: SARRIA-MILLAN, RODRIGO
TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
TITLE OF INVENTION: IN PLANTS
FILE REFERENCE: 16313-0031
CURRENT APPLICATION NUMBER: US/10/768,511
CURRENT FILING DATE: 2004-01-30
PRIOR APPLICATION NUMBER: 60/196,001
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 723
TYPE: DNA
ORGANISM: Physcomitrella patens
US-10-768-511-2

Query Match 15.5%; Score 673; DB 7; Length 723;
Best Local Similarity 97.1%; Pred. No. 2.3e-200; Indels 1; Gaps 1;
Matches 696; Conservative 0; Mismatches 20;

QY 2753 GAAGCTCGAGATCGTATTGAGGTAAACAGGAGTTTTCAGGCGCATGGCAGTTTCAGATTGG 2812
DB 3 GCACAGGAGATCGTATTGAGGTAAACAGGAGTTTTCAGGCGCATGGCAGTTTCAGATTGG 62

QY 2813 TCCGAATCAACGAACTATTCAGGACATTTGTAAGAACCCTACATCGATTCCGTGCGAGTCAA 2872
DB 63 TCCGAATCAACGAACTATTCAGGACATTTGTAAGAACCCTACATCGATTCCGTGCGAGTCAA 122

QY 2873 GAAGTCTGACAGGGTTCAGTCTCAAACTGAAGATCCTATGGAGTGGTAAGGAGAATGA 2932
DB 123 GAAGTCTGACAGGGTTCAGTCTCAAACTGAAGATCCTATGGAGTGGTAAGGAGAATGA 182

QY 2933 TATGTATGCTGGATCATGAAGTGATCTTCAAGAGCTGCTAATGAAGCAAAAGATTCA 2992
DB 183 TATGTATGCTGGATCATGAAGTGATCTTCAAGAGCTGCTAATGAAGCAAAAGATTCA 242

QY 2993 AAAAATTAAGAGCTGTCTCAAGCTCCCGGCAATTTATGATAGATCTTCAAGGTCGCTGGC 3052
DB 243 AAAAATTAAGAGCTGTCTCAAGCTCCCGGCAATTTATGATAGATCTTCAAGGTCGCTGGC 302

QY 3053 TCAAGCATTTGGAGCTTGAAGATATTAAGAGGCTCTTCTGCGAGCTCTTTGGTGG 3112
DB 303 TCAAGCATTTGGAGCTTGAAGATATTAAGAGGCTCTTCTGCGAGCTCTTTGGTGG 362

QY 3113 GAAGGCTAAGAAATTCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTACTTGT 3172
DB 363 GAAGGCTAAGAAATTCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTACTTGT 422

QY 3173 TGGGGACCTTGTACAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCTCTCG 3232
DB 423 TGGGGACCTTGTACAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCTCTCG 482

QY 3233 TGAATCTACATAGTGGGCGAGAGTTTCGGGGTTCGGCTGACAGCGTATCT-AACCA 3291
DB 483 TGAATCTACATAGTGGGCGAGAGTTTCGGGGTTCGGCTGACAGCGTATCTAAGCGA 542

QY 3292 AGGATCCAGAACTCCAGAGACCGTATTTGGAGAGCGAGCTTTGGTCTTCTAGTATCGTG 3351
DB 543 AGGATCCAGAACTCCAGAGACCGTATTTGGAGAGCGAGCTTTGGTCTTCTAGTATCGTG 602

QY 3352 GGTATGCTGTATCGATGAGTTTCAGCAAAAATGTTGATAATGCCCGAGCATGCTTCATG 3411
DB 603 GGTATGCTGTATCGATGAGTTTCAGCAAAAATGTTGATAATGCCCGAGCATGCTTCATG 662

QY 3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCAAGGGGGTATCATTTGCCCTCGCTGA 3468
DB 663 AGGTAATGGAGCAACAAACGGTATCTGTAGCAAGGGGGTATCATTTGCCCTCGCTGA 719

US-10-425-114-19722
Sequence 19722, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 19722
LENGTH: 1571
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3150-112-C5_FLI
US-10-425-114-19722

Query Match 14.6%; Score 634.2; DB 7; Length 1571;
Best Local Similarity 67.5%; Pred. No. 6.6e-188; Indels 3; Gaps 1;
Matches 907; Conservative 0; Mismatches 433;

QY 2917 TGGATAGGAGATGATATGTGCTGGGTATCATGAAGTGATGATCTTCAGAGCTGCTA 2976
DB 32 TGGAGGACACCATGGATTTGATTAATTTCTACGCTGACAAATCTACTGAAGAGGATTTTC 91

QY 2977 ATGAAGCAAAAGATTCAAAAACCTTAAAGAGCTGTCCAAAGCTCCCGGCAATTTATGATAGAC 3036
DB 92 TTAGTGATAGGTTGAGAACTTAAAGAGCTTTCCAAAGTTGCTGATATCTATGAAGAT 151

QY 3037 TTTCAAGGTCGCTGGCTCCAAAGCAATTTGGGAGCTTTGAAGATTTAAAGAGGCTTCTCTT 3096
DB 152 TGACTAGATCATTAAGCTCCAAACATATGGGAGTTGGATGATGTCAAAAGAGGCTCTCTT 211

QY 3097 GCCAGCTCTTTGGTGGGAGGCTAAGAAAATTCATCTGGAGCATCTTTCCGAGGTGACA 3156
DB 212 GCCAGCTCTTTGGGCGCAATCCCTTGAAGCTTCTCTGGAGCTAGTTTCCGGGCTGACA 271

QY 3157 TCAATGTTTTACTTTGTTGGGACCCCTGTAACAGTAAATCTCAGCTGCTCAGTATGTC 3216
DB 272 TCAATATTTTACTTTGTTGGGACCCCTGGAACAAAGTAAATCCAGCTTCTCCAGTACATGC 331

QY 3217 ACAAGATAGCTCTCTGCGAATCTACACTAGTGGGCGGAGGAGTTTCGGGCTGCGCTGA 3276
DB 332 ATAAACTGTCTCTCTGCTGATCTATACGAGTGGTAGAGAAAGTTCTGCTGTTGCTTTA 391

QY 3277 CAGCGTATCAAGAGGATCCAGAAATCTCGAGAGACGGTATTTGGAGAGCGGAGCTTTGG 3336
DB 392 CTGCTTATGTTACCAAGACCTGAGACTGGCGAACTGTTCTAGAAAGCTGGAGCACTTG 451

QY 3337 TTTCTAGTATCGTGGGATATCTGTATCGATGAGTTTCAGAAAATGCTGATAATGCC 3396
DB 452 TTTTGGTGCACAAAGGTTGTTGTTGGTAGATGAGTTTGTATAAGATGCTGATAATGCC 511

QY 3397 GAAGCATGTTTATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCA 3456
DB 512 GAAGCATGTTTACAGAGGTGATGGAACAGCAGACATATCCATTTGCGAAGGCTGGGAATA 571

QY 3457 TTGCTCTGCTGAGACGCTCCGAGCTGTCTCTGATGTGCAATCTCTAGTGGGCTCCCGAT 3516
DB 572 TTGCTCTTTTAAACCGCTAGACATCTGTCTGCAATGTGCCATCTCTACTGAATCAGTT 631

QY 3517 ACAATCGCGGCTTTCTGTGATTGATAACATCAGCTTCTCTCAACTCTACTCTTTAGAT 3576
DB 632 ACAATCCAGGCTCTCTGTAAATTTGACAAACATCCACTTAGCGCCAAACGCTACTTTCAAGAT 691

QY 3577 TTGATTTAATTTACTTAATGCTCGACAAACAGCAAGCAAAACGATCGTCTCGCCA 3636

Db 692 TCACCTGATTTATCTTATCTTGGCAAGCGGATGAGCAAACTGATAGCGCTTGGCAA 751
QY 3637 GGCATCTCGTGGCTTTACACTATGAAAACTATCAAGTTTCAAAGCAGGAGCGCTTAGATC 3696
Db 752 AGCATATGTTGTTGCTTTGAGNAATCCAAATTTAGAGGAGCTCGAGGCTTTGGACT 811
QY 3697 TACAAACACTTACCGGTATATCACCTATGCTCGTCAGCAATGATACCTTACATTAAGTG 3756
Db 812 TGCAGACACTAGTATTTCTACATAAGCTATGCAAGGAATATATTCAGCCACAGATTATCTG 871
QY 3757 ATCAAGCTCTGAAGATTGATTAATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTC 3816
Db 872 ATGAAGCTGCAGAAGAGTTAACTCGTGGCTATGTCGAGATGAGAAAAGAGGGAATAGCC 931
QY 3817 CTGGAAGCAGTAAAGGTGATAACAGCCACACTCGGCAACTCGAAAGTATGATTCGTA 3876
Db 932 CTGGAGCAGAAAGAGGTATTAACAGCAACCCCTAGACAAATAGAGAGTTGATTCGTC 991
QY 3877 TCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTGAAGAAAGTTGATGCAAGCAG 3936
Db 992 TCAGCGAAGCATTAGCCCGAATGCGGTTCTCTGAAGTGGTTCGAGGTGCGGGAATGTTGG 1051
QY 3937 AAGCTGCGCCCTTTAGAGCTGCTTTGAGCAATCTGCTACTGATCATGCAACAGGTA 3996
Db 1052 AGCATTTCAAGCTTCTTGAAGTCGCCATGCGCAGTCTGCGACGGATCATGCAACTGGTA 1111
QY 3997 CGATAGACATGATCTTATCAACACTGGAGTTCGGCCAGCGAGCGTATTCGTCGGGCCA 4056
Db 1112 CGATTGATATGATCTGATCATGACGGGATATCCGACGAAAGCGAGGCGGGAGA 1171
QY 4057 ACTTGTAGCTGCTCTGCGAGACTTATAGCAGATAAAATTTCACTGGCAGCTCTCTTG 4116
Db 1172 ACCTGCTGGCGCAACCCGTAACCTGATTCGGGAGAAATGACGCTTGGAGGCCCTC-- 1229
QY 4117 GCTTGAAGCACTGAGCTTCTTGAAGATATCCGAGCCAAAGCAGTGTGGAGCTTAGTT 4176
Db 1230 -GATGCGCATGATGATGTTGCTGCGAGGAATGAGGAAGCAGAGCTCAATGGAATTCATA 1288
QY 4177 TGCAGGATATTAATAATGCTCTGGTAGCTTCCAGGAGAGGCTTCTTACTGTCCATG 4236
Db 1289 TGCAGCACTCCCGGCTGCTCTTGGCACTGATGATGAGGAAGCGGCTGATTCATG 1348
QY 4237 GTGACATAGTCAAGAGATTGA 4259
Db 1349 GAGCAACGTCGAGGAGATTGA 1371

RESULT 9

US-10-425-115-98897
; Sequence 98897, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98897
; LENGTH: 2939
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21707C.1
US-10-425-115-98897

Query Match 13.8%; Score 598.6; DB 8; Length 2939;
Best Local Similarity 56.7%; Pred. No. 1.7e-176;

Matches 1158; Conservative 0; Mismatches 864; Indels 21; Gaps 2;
QY 2217 GAGGATCTCTAAATATCGACATGTGACACATTTATGACCATGATCCTGATCTATACGCA 2276
Db 638 GGGAGTCGCTCGAGTCGATGCGCAGCAGCGTGTGACCAAGACCCAGACCTCTACAGC 697
QY 2277 AAAATTGTTGATACCCACTCGACATCATCCCTCTTGGACACTGAGTGTGAGGAAT 2336
Db 698 AAGATGTTGCTATCCGCTCGAGGTGCTCGCATCTTCGACATCGTGTCTCATGGACCTC 757
QY 2337 GCTACCTCTTTACTACCAACGTTTGAAGACATATTGAGCCAGACCTTTCAATCTCAAA 2396
Db 758 GTGCGCGCATCGAGCCGCTCTTCGAGAAGCAGATCCAGACCAAGGATCTTACAACCTCAAG 817
QY 2397 GCATCGGTGCACATCGTGAACCTCAACCCCTTCAGATATAGACAAATTTGGTTTCTGTAAA 2456
Db 818 TCGTCCATTTGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAG 877
QY 2457 GGAATGGTTATCCGGTGCAGTTCATCATACCTGAAATTAAGGGGGCTTCTTCAATGT 2516
Db 878 GGTATGATTAATTAGATGCACTCGGTCTATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGC 937
QY 2517 TTAGTGTGTGGTCACTCGCTCGCTAGTTTACAGTTTGTAAAGGGCGGGTTGAGGAGCCA 2576
Db 938 CTGCTTTGTGGTTCCTCTCAGAGCCCGTCTATGTTGATAGAGAAGAGTAACCTGAACCA 997
QY 2577 ACAAGGTGAAAAGCCAGAAATGTGACGACCGGAATGCTATGTCTCTTATTCAAATCGA 2636
Db 998 CACATTTGTGAGAAAGCAATGTAAGCCACAAATTTCTATGACCTCTAGTGCACACAGA 1057
QY 2637 TGCATTTTGCATAATGACAGATAGTGGCTCTTCAAGAAATCTCCAGATGCCATTCCTGAA 2696
Db 1058 TGCAGATTTTTCAGCAAGCAGATCATAAAGTTGCAAGAAACACACAGACGAGATACCAGA 1117
QY 2697 GAGAGACTCCACACACAGTACGATGTCTTTATACAACACTATGTTGATGCTGTGTGAAG 2756
Db 1118 GGTGGCACTCCACATACAGTTAGTGTCTTGTGATGATGAAGCTTGTGATGCTGGAAG 1177
QY 2757 CTTGAGATCGTATTGAGGTAACAGGAGTTTCAAGGCCATGCGAGTTCGAGTTGTCG 2816
Db 1178 CTTGAGATAGGTTGAGATTAATGGAATATACAGAGCTATGAGTATTCGAATTTGACCA 1237
QY 2817 AATCAAAGCAATTAACGACATTTGATAAGACTCATCATGTTGCTGCAAGTCAAGAG 2876
Db 1238 ACTCAAAGGACAGTGAAGTCTATATTCAAGACATATATTGCCGGATGCGGTTCTCTGAA 1297
QY 2877 TCTGACAGGGTGCAGTGCAAACTGAGATCCTATGAGATGATTAAGGAGATGATG 2936
Db 1298 GTGGTCGAGGTGCGAGATGTTGTGGAGGCTTCAGGCTTCTTGAAGTTGCCATGCGAGCAG 1357
QY 2937 TATGCTGGGTATCATGAAAGTGATATCTTCAGAGCTGTCTAAATGAAGCAAGATTCAAAA 2996
Db 1358 TCTGCAACCGATCATGCACTGGGAGGATGATATGATCTAATCATGACGGGATATCT 1417
QY 2997 CTTAAAGAGCTGTCCAAGCTCCCGGCAATTTATGATAGACTTTTCAAGGTCGCTGGCTCCA 3056
Db 1418 GCAAGTGAAGGCGAGAGACGGAACGACCTTGTGTCT-----GCAACC 1459
QY 3057 AGCATTTGGAGGCTTGAAGATTAATAAAGGGTCTCTTTGCCAGCTCTTTGGTGGGAAG 3116
Db 1460 CGTAACCTGATCGTGGAGAAAATGACGCTCGGAGGGCTTCAATGATGATGATGATG 1519
QY 3117 GCTAAGAAAATTCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGTGGG 3176
Db 1520 CTGTAAGAGCTCAGTTCGGGAGCTGAATGGAGAGCCATCTGCACGAATTTCTTCGGGGCG 1579
QY 3177 GACCTCTGTACCAAGTAAATCTTCAGCTGCTTTCAGTATGTGCAACAAGATAGCTCTCTGTGA 3236
Db 1580 CTTGGCACTATGATGACAGAGGTGTGCTTCTGATCCATGGAATAAAGGTGAAGAGAGGT 1639
QY 3237 ATCTACACTAGTGGGCGAGGAAGTTCCGGGGTTGGGCTGACAGCGTATGTAACAGAGGAT 3296
Db 1640 TGAATTTAACTGCCATCTGCAGGGAATGGTTGTGTGAGCTCCGCTCTACCGTTTCTTCAAC 1699

3297 CCAGAACTCGAGACGGTATTGGAGAGCGAGCTTTGGTTCTTCTAGTGATCGTGGGATA 3356
Db
1700 AGTCAAGGTTGCGAAGATCTGATCGACTGTAGGGCAATTTGCTTCAGAGTGTAAAGTAGT 1759
Qy
3357 TGCTGTATCGATAGTTTCGACAAAATGCTGTGATTAATGCCCAGAGCATGCTTCATGAGGTA 3416
Db
1760 TTGTCTCGATGAGTTTTCATAAAGATGTCAAGTAAATGCCCAGAGCATGTTACAGAGGTG 1819
Qy
3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTTGCTCGCTGAACGCTCGG 3476
Db
1820 ATGGAACAGCAGACATATCCATTGCGAAGGTGGAATAATTCGATCTTTAAACGCTAGG 1879
Qy
3477 ACGTCTGCTTGTGATGTGCAATCCTAGTGGGTCGCCGATACAATGCGCGCTTTCTGTG 3536
Db
1880 ACACTGCTCTGGCATGTGCCATCTTACTGATCAGGTACAGTACAAATCCAAAGGCTCTCTGA 1939
Qy
3537 ATTGATAACATCCAGCTTCTCTCAACTCTCTACTCTCTAGATTTGATTTAATTTACTTAATG 3596
Db
1940 ATTGACAACATCCACTTAGCGCAACGCTACTTTCAAGATTGCACTGATTTATCTTATC 1999
Qy
3597 CTCGACAAACAGACAGCAACAGATCGTCTGCTCGCCAGGCATCTCGTGGCTTTACAC 3656
Db
2000 TTGGCAAGGCGGATGAGCAAACTGATAGGCGCCTGGCAAGCATATTGTTTCGTGTCAT 2059
Qy
3657 TATGAAACTATCAAGTTTCAAAAGCAGGACGCTTTAGATCTACAAACACTTACCGCGTAT 3716
Db
2060 TTGAGATCCAAATTTAGAGAGCTCGAGGTCTTTGGACTTGCAGACACTAGTTTCTTAC 2119
Qy
3717 ATCACTATGCTGCTCAGCATGTACATCTTAATTAAGTGATGAAGCTGCTGAAGATTG 3776
Db
2120 ATAAGCTATGCAAGGAAGTATATTACGCCACAGTTATCTGATGAAGCTCGAGAAGATTA 2179
Qy
3777 ATTAATGGCTATGTTGAGATGCGCAAAAGGCACTTTCTCGGAGCAGTAAAGGTG 3836
Db
2180 ACTCGTGGCTATGTGAGATGAGAAAGAGGGATAGCCCTGGGAGCAGAAAGGTGTC 2239
Qy
3837 ATAAACAGCACACCTCGGCAACTCGAAAGTATGATTGCTATCAGTGAAGCCCTAGCTCGA 3896
Db
2240 ATAAACAGCAACCGCTAGACAAATAGAGAGTTTGTATCCGCTCAGCGAGCATTTAGCCGA 2299
Qy
3897 ATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGACAGCAAGCTGTGCGCTTTTGAAC 3956
Db
2300 ATGCGGTCTCTGAAGTGGTTCGAGGTGCGGATGTTGTGAGGCAATTCAGGCTTCTTGA 2359
Qy
3957 GTGCTTTGAGCAATCTGCTACTGATCATGCAAGGATGAGTACGATAGCATGATGATCTTATC 4016
Db
2360 GTCGCCATGACGAGTCTGCGACGGATCATGCAACTGGTACGATGATATGGAATCTGATC 2419
Qy
4017 ACCACTGGAGTGTGCGCCAGCGAGCGTATTCTGTCGGGCCAACTTGTCTAGCTGCTTCGCA 4076
Db
2420 ATGACGGGNTATCCGCAAGCGAAAGCGAGAGCGGGAGNACCTCGTTGCGCAACCCGT 2479
Qy
4077 GAGCTTATAGCAGATAAAATTTCACTGGCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4136
Db
2480 AACCTGATTGCGGAGAAATGCACTTGGAGGCCCTC---GATGCGCATGATTGAGTTG 2536
Qy
4137 CTTGAGGATATCGGAGCCHAAAGCAGTGTGAGGTTAGTTTGGCAGGNTATTAATAATGCT 4196
Db
2537 CTGGAGGAACCTGAGGAGCAGAGCTCAATGGAAATTCATATGACGCAACTCCCGCGTCT 2596
Qy
4197 CTGGGTAGCTTCAAGGAGAGGCTTTCTTACTGTCTCCATGCTGATGATGATGATGATGATGATGAT 4256
Db
2597 CTTGGCACCTTGATGACTGAAGGCGCGTGGTTATCCATGAGACAACTGAGGAGNTT 2656
Qy
4257 TGA 4259
Db
2657 TGA 2659

; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2332
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
; US-10-320-797-2332

Query Match 11.3%; Score 492.8; DB 7; Length 2970;
Best Local Similarity 57.1%; Pred. No. 3.8e-143;
Matches 993; Conservative 0; Mismatches 717; Indels 30; Gaps 4;

Qy 2372 TGAGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAAACCTTCAGA 2431
Db 768 TAAAGTCAGRCGTTTGGTGAGAGAACGCGTCAACATGAGGATCTTAATCCCGAGA 827
Qy 2432 TATAGACAAATTTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCACTTCTATCATACCTGA 2491
Db 828 TACTGACAACTGGTGACCGTGAAGGTCCTTGTCTATCCGTCGCAACTCCGCTCATTCCTGA 887
Qy 2492 AATTAAGGGGCGCTTCTTCAAATGTTAGTGTGGTCACTCGCTCCGCTAGTTACAGT 2551
Db 888 TATGACCACTGCTTCTTCCGATGCTTGGTTGTGAGCACACTGTTCAAGCGGATA---T 944
Qy 2552 TGTTHAAGGGCGGTTGAGGAGCCCAACAGGTGTGAAAGCCAGAAATGTGCAGCAGGAA 2611
Db 945 CGACCGAGAGAAATCAGCGAGCCAGAGCGATGTCCCGAGACGCTCTGTGGTTCACTGG 1004
Qy 2612 TGCTATGCTCTTATTTCAAAATCGATGCATTTTGCAAATAGCAGATAGTGCCTTCA 2671
Db 1005 CACTATGCTCTTATTTCAATACCGCTCCGAATTCACCGCAACAAAGTCACTCGCTTCA 1064
Qy 2672 AGAACTCCAGATGCCATTTCTGAAGGAGAGACTCCACACAGTCAGCATGTGTTTATA 2731
Db 1065 GGAGACTCCGATGCTGTACCCGATGGGCAGACTCCACATACTGTATCTCTTTGGCTCTA 1124
Qy 2732 CAACTATGTTGATGCTGTGAAGCCTGAGATCGTATTGAGGTAACAGGAGTTTCAA 2791
Db 1125 TGATGAACCTCGTTGACTTTGGTCAAGCCTGGTGACCGAGTGATCATCCTGGTATCTCG 1184
Qy 2792 GGCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTTACGAGCATTTGTATAAGACCTA 2851
Db 1185 ATCAATCCAGTTCGTGTCAACCCCTCGCCAAAGGATATCAATCTTTGTACAGACCTA 1244
Qy 2852 CATCGATTGCTGCAAGTCTCAAGAGTCTGACAGGGGTGCACTGCAAACTGAAAGATCCTAT 2911
Db 1245 TCTCGATGCTGCCAGTCAAGCGTACCAACACTGCCCGAATGGGCTTTGACCTTCTA- 1303
Qy 2912 GGAGATGGAATAGGAGATGATATGTATGCTGGGTATCATGAAAGTGATCTTCAGAGC 2971
Db 1304 -----CTCGTGGCGGAGAGCAAGCCACCGGTGACCTTTGGGTGCGCTCTCAGATC 1355
Qy 2972 TGCTAATGAAGCAAGATTTCAAAACTTAAAGAGCTGTCCAAAGCTCCCGGGCAATTTATGA 3031
Db 1356 AGCCGCTGGGAGATGGAACACGCAATTCGAGCTCTCCAAACCCTGATCTTTACAA 1415
Qy 3032 TAGACTTTCAAGTCTGCTCCAAAGCATTTTGGGAGCTTTGAAGATATTTAAAGGGTCT 3091
Db 1416 CATACTTGCTCATCTTTGGCGCCCTCCATCTACGAGCTCGAAGATGTTTAAAGGGCAT 1475
Qy 3092 TCTTTCGACGCTCTTTGGTGGGAGGCTAAGAAATTCATCTCGAG----- 3138

Db 1476 CCTCTCCAGCTTTTGGCGGTATCCAAAGTCTATGCGCCGTTGGTGGCGGTGGGGCGG 1535
 Qy 3139 --CATCTTCCGAGGTGACATCAATGTTTACTTGTGTTGGGACCCCTGGTACCAAGTAAATC 3196
 Db 1536 TCCTAGATACAGAGGTGATCAACGTTGTTGATGTTGGTATCCTGGTACTAGTAAATC 1595
 Qy 3197 TCAGCTGCTTCAGTATGTGCAAGATAGTCTCTCTGTTGGAATCTACACTAGTGGCGAGG 3256
 Db 1596 TCAGATCTGCAATGTGATGAATGCTCTCTGAGGTGTCTACACCTCCCGTAAAGG 1655
 Qy 3257 AAGTTGGCGGTTGGCTGACAGGTATGTAAAGAGTCCAGAACTCCAGAGCGGT 3316
 Db 1656 AAGTTGAGCGGTTGGTTGACGCGCTATGTAAAGAGATCCGATTCAGAGCGTCTGT 1715
 Qy 3317 ATTGGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTGCA 3376
 Db 1716 TTTGGAAAGTGTGCTATGTTTGTCTGACGGGAGTTTGTGTATCGATGAGTTTGA 1775
 Qy 3377 CAAATGCTGTGATAATGCCGAGAGTCTTCTCATGAGGTAAATGAGCAACAAACGGTATC 3436
 Db 1776 CAAGATGAGTGTATGCCACTCCAAAGTGTCTTGCACGAAGTAAATGAGCAACAAACGGTTTC 1835
 Qy 3437 TGTAGCAAAAGGGGTATCATTCCTCGCTGAAGCTCGGACGCTCTGCTTCATGTGC 3496
 Db 1836 TATTGCCAAGCGAGGTATCATCACCACTCTCAATGCTCGTACATCCATTCGCGCGCGC 1895
 Qy 3497 AATCTCTAGTGGGTCCGACATCAATGCGCGCTTTCTGTGATGATTAACATTCAGCTTCC 3556
 Db 1896 GNAACCAATCACTCTGATGATGACCTTACCTCCCTATTCCTGCGCAAGATCGATCTCC 1955
 Qy 3557 TCCAACTCTACTTCTAGATTTGATTTAAATTTACTTAAATGCTCGACAAACGACGAGCA 3616
 Db 1956 TCCTACATTTGATCTCAAGATTGCAATGCTGCTGATCTGCTGTTGACAAAGGTGATGAGT 2015
 Qy 3617 AAGCATGCTGCTCTGCGGAGGATCTGCTGGCTTTACATATGAAACTATGAACTTTC 3676
 Db 2016 GNAACGACAAAGTTGCAAGCAATTTGGTGGGACTCTTATTTGTGATGAGGATCA 2075
 Qy 3677 AAGCAGGAC---GCCCTTAGATCTACAAACACTTACCGCTATATCACCTATGCTGCTCA 3733
 Db 2076 GCCTGCGCAATATCATCTCTCTTACAGCTTAAAGCTCTACATTAATATGCGCGCTC 2135
 Qy 3734 GCATGTACATCTTACATTAAGTATGATGAGTCTGTAAGATTTGATTAATGGCTATGTTGA 3793
 Db 2136 CAAATCCACCCGCTCTCCAGGAGGCTTCTGAGGCGCTTGTTCAGGCTATGTTGA 2195
 Qy 3794 GATGCGCAAAAGGCAACTTTCTCGAAGAGTAAAGGTGATTAACAGCACACCTCG 3853
 Db 2196 GATGCGCAAGCGGTATGGAATCTCTGATCTCAGGAAAGAGGATCACTGCTACCACTAG 2255
 Qy 3854 GCACTCGAAGTATGATTCGATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGT 3913
 Db 2256 GCAGTTGGAAGTATGATCAGATTGGCGAGCGCATGCGAGGATGAGATTAAGTGACAG 2315
 Qy 3914 GGTAGAAAGATTGATCAGCAGAGTGTGCGCTTTTATAGCGTCTGTTGACGCAATC 3973
 Db 2316 GGTGAGGAGAGGACATTAGGAGGCTGTTAGTTGATCAAGAGTGCCTGAGGAGAG 2375
 Qy 3974 TGCTACTGATCATGCAACAGGTACGATGATGATCTTATACGATGAGTGTGCGG 4033
 Db 2376 TGCTACCGATCCCTTACTGCTCAAAATGATCTTGAACCTCATCAACACCGGTGCGGCCA 2435
 Qy 4034 CAGGAGGATTCGTTGGGCAACTTGTAGCTGCTCTGCGAGGCTTATAGCAGATAA 4093
 Db 2436 AACCATCGTGGCGCGCGGCGGATCTTAAACGTGAAGTTATCAAACTCGTCTCGAAAA 2495

RESULT 11

US-10-260-238-580
 ; Sequence 580, Application US/10260238
 ; Publication No. US20040016025A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.
 ; APPLICANT: Briggs, Steven P.
 ; APPLICANT: Cooper, Bret
 ; APPLICANT: Glazebrook, Jane
 ; APPLICANT: Goff, Stephen A.
 ; APPLICANT: Katagiri, Rumiaki
 ; APPLICANT: Krepis, Joel
 ; APPLICANT: Provart, Nicholas
 ; APPLICANT: Ricke, Darrell
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
 ; FILE REFERENCE: 60111-NP
 ; CURRENT APPLICATION NUMBER: US/10/260,238
 ; CURRENT FILING DATE: 2002-09-26
 ; PRIOR APPLICATION NUMBER: US 60/325,448
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: US 60/325,277
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: US 60/370,620
 ; PRIOR FILING DATE: 2002-04-04
 ; NUMBER OF SEQ ID NOS: 6077
 ; SEQ ID NO 580
 ; LENGTH: 1723
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; NAME/KEY: N region
 ; LOCATION: (95)..(95)
 ; OTHER INFORMATION: n = any nucleotide
 US-10-260-238-580

Query Match 11.3%; Score 490; DB 7; Length 1723;

Best Local Similarity 69.4%; Pred. No. 2e-142;

Matches 762; Conservative 0; Mismatches 285; Indels 51; Gaps 5;

Qy 2989 TTCAAAACTTAAGAGCTGTCCAAGCTCCGGGCATTTATGATAGACTTTCAAGTCCG 3048
 Db 606 TTCACATAAAGAGGAGTTGTCAAGTTGCCAGATATATGACAGATTAACTAGGTCA 665
 Qy 3049 TGGCTCCAAGCATTTGGGAGCTTTGAAGATATTTAAAGAGGCTTCTTTTC---CAGCTCT 3105
 Db 666 TGGCTCCAACATATGGAGCTGCGACGAGTTTAAAGGGGCTCTTTTGCAGCAGGATTT 725
 Qy 3106 TTGGTGGAGGCTAAGAAATTCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTT 3165
 Db 726 TTGGTGGAAATGCTTTGAGGCTTCTTCTGAGCTAGTTTCCGAGGCGCATCAATATTT 785
 Qy 3166 TACTTGTGGGACCTTGTTACAGTAAATCTCAGTGTCTCAGTATGTCGACAGATAG 3225
 Db 786 TGGCTTGTGTGATCTCTGGAACAAGTAAATCCAGCTTCTCCAATACATGACAACTGT 845
 Qy 3226 CTCCTCGTGAATCTACACTAGTGGGCGAGGAAGTTTCGGGCTGACACGCTATG 3285
 Db 846 CTCCTCGTGCAATTTATCAAGTTGGCAGAGGATTCAGCTGTTGGCTTACTGCAATACG 905
 Qy 3286 TAAACGAAGGATCCAGAACTCGAGAGACGGTATTTGGAGAGCGGAGCTTTGGTCTTAGTG 3345
 Db 906 TTACCAAGGATCTGAACTGGTGAATCTTCTTGAGAGTGGAGCGCTTTGTTTGTAGTG 965
 Qy 3346 ATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATTAATGCCGAGGATGC 3405
 Db 966 ACAAGGTTGTTGTTGATGAAATTTGATGAAGATGTCTGATAATGCTCGAAGCATGT 1025
 Qy 3406 TTCATGAGTAAATGAGACAAACGGTATCTGTAGCCAAAGGGGTATCATTCGCTCCG 3465
 Db 1026 TACATGAGGTGATGGAAACAACAGACTGTCTCCATTTGCCAAGGCTGGAAATTAATTCATCTT 1085
 Qy 3466 TGAACGCTCGGAGCTGTGCTCTGATGTGCAATCTTAGTGGGTCCCGGATCAATGACGCG 3525
 Db 1086 TAAATGCCAGAAATCAATGATTTCTAGCATGTGCAATCTTACTGATCAGCTTATATCA 1145
 Qy 3526 GCCTTTCTGTGATGATTAACATCCAGCTTCTCCAACTCTACTTCTTCTAGTA--TTTGTATTT 3583

```
Db 1146 GGCTCTCTGTGATTGACAAATATCCATCTTCTCCAAACACTGCTTTCTAGAGGTTTGACCT 1205
Qy 3584 AATTTACTTAATCTCGACAAACAGCAGCAAAACGATCGTCTCGCCAGGATCT 3643
Db 1206 CATTTATCTGATATTGCAAGGCAAGATGAGCAAACTGATAGACGCTCGCTGAAGCATAT 1265
Qy 3644 CGTGGCTTTACACTATGAAGAACTATGAAGTTTCAAAGCAGGACGCCCTTAGATCTACAAAC 3703
Db 1266 TGTTTCGTTGCAATTTGAGAA----- 1286
Qy 3704 ACTTACCGGCTATATCACCTATGCTCGTCAGCATGTACATCTCTACATAAAGTGATGAAGC 3763
Db 1287 ---TCCAAACTACATAAGTTATGCAAGGAAGCATATACAAACACACAGTTATCTGATGAAGC 1343
Qy 3764 TCGTGAAGATTTGATTAATGCGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTCTGGAAG 3823
Db 1344 TGCAGAAGAAATTGACTCGCGGCTATGTTGAGATGAGGAAAGAGGAAACAGCGCCCTGGTAG 1403
Qy 3824 CAGTAAAGAGGTGATAACAGCCACACTCGGCAACTCGAAAGTATGATTCGTATCAGTGA 3883
Db 1404 CAG---AAGAAGATAACTGCGACAGCTCGACAAATGAGAGCTTGATTCGGCTCAGTGA 1460
Qy 3884 AGCCCTAGCTCGAATGAGATTTTCTGAAGTGTGTAGAGAAAGTTGATGCGACGAAAGCTGT 3943
Db 1461 AGCACTGGCCGGAATCGGATTTCTCGAATGTTGGAAGTACAAAGATGTTGTAGAGGCCCTT 1520
Qy 3944 GCGCCTTTTAGAGCTGCTTTGAGCAATCTGCTACTGATCATGCAAC-AGGTACGATAG 4002
Db 1521 CAGGCTTCTCGAAGTTGCCATGAGCAATCGGCNACTGATCATGCCACTGGGTACAAATCG 1580
Qy 4003 ACATGATCTTATCAGCACTGGAAGTGTGCGCCAGCGAGCGTATTTCTGTCGGGCCAACTTGC 4062
Db 1581 ATATGATCTTATCATGACTTGAATATCTGCGAGCGAAAGGCGAGCGCGACCAAGTTGC 1640
Qy 4063 TAGCTGCTCTGCGAGAGC 4080
Db 1641 TCGAAGAAATTAGGAGC 1658

RESULT 12
US-10-032-585-6480
; Sequence 6480, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6480
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6480

Query Match 11.0%; Score 480.2; DB 6; Length 2733;
Best Local Similarity 56.6%; Pred. No. 3.4e-139;
Matches 959; Conservative 0; Mismatches 718; Indels 18; Gaps 3;

Qy 2409 ATCGGTGAACCTCAACCCCTTCAGATATAGACAAATTTGGTTTCTGTAAAGGAATGGTTATC 2468
Db 883 ATTCTGTGAATTTGATCCTTAATGATATTGATAAGTTAGTTAGTGTAAAGGGTTAAACATTA 942
Qy 2469 CGGTGAGTTCTATCATCTCTGAAATTAAGGGGGCCTTTCTCAAAATGTTTAGTGTGTGGT 2528
Db 943 CGTTCTACTTCTATCATCTCTGATATGAAAGTGGCATTTTTCAGATGTAAACGCGCTGTGGT 1002
Qy 2529 CACTCGCCTCCGCTAGTTTACAGTTGTTAAAGGGGGGTTGAGGAGGCAACAAAGGTGTGAA 2588
```

```
Db 1003 CATACTGTTGGG---GTGGAAATTTGATCGTGTGTAAATTTCCGAACCAACAAATGTCTCT 1059
Qy 2589 AAGCCAGAAATGTGACAGCAGGAAATGCTATGCTCTTATTTCAACAATCGATGCATTTTGCA 2648
Db 1060 AGAGAAGTATGTGGTCAAAACCAATCAATGCTGTTATTTTCAATAATGCTTATCATCTTTCT 1119
Qy 2649 AATAAGCAGATAGTGGTCTTCAAGAACTCCAGATGCCATTCCTCAAGGAGAGACTCCA 2708
Db 1120 GATAAACAAGTGATTAATAATTAAGAAACTCTGATTTGGTACCTGATGTTGTTCAACCCCT 1179
Qy 2709 CACACAGTCAGCATGTGTTTATACAACTATGTTGTTGATGCTGTGAAAGCTTCGAGATCGT 2768
Db 1180 CATCTATCAATTTAATGTTTACGATGAATTTGTTGATTTCTTGTGCTGCTGATCGT 1239
Qy 2769 ATTGAGGTAAACAGGAGTTTTCAGGCCATGCGCAGTTTCGAGTTGCTCGGAATCAACGAACA 2828
Db 1240 GTTGAAGTATGTGGTATTTTCAGATCAACTCCAGTTCGAGCAAACTCTAGACAAAGAGCT 1299
Qy 2829 TTACGAGCATTCGTATAAGACCTACATCGATTGGTGCACGTCAAGAAAGTCTGACAGGGGT 2888
Db 1300 TTAAAAAATCTTATATAAAAAATTTTGGATATTTGTTGATGAAAAAATTCGATATAAAGG 1359
Qy 2889 CGACTGCAAACTGAAAGATCCTATGAGATGGATTAAGGAAATGATATGTATGCTGGGTAT 2948
Db 1360 CGATTAGGTGGTGTATTACCA-----CTTTAGAACACGAATTTGGCTGAAAAAGGAT 1410
Qy 2949 CATGAAAGTGATACCTTCAGAACTGCTAATGAAGCAAAAGATTCAAAAATTTAAAGAGCTG 3008
Db 1411 CAAGAAGTTGAACAAGTTTCGAAAAATCACTGCTGAAAGAGAAGCTAAAAATCAAGAATAAT 1470
Qy 3009 TCCAAAGCTCCCGGCATTTATGATAGACTTTCAAGTCTGCTGCTCCAGCAATTTGGGAG 3068
Db 1471 TCTGAACGTGATGATTTGATGAAATATTTGGCTAGATCAATTAGCTCCATCTATTTATGAG 1530
Qy 3069 CTGGAAGATATTAAGAGGCTCTTCTTCCAGCTCTTTGGTGGGAAGGCTAAGAAAAAT 3128
Db 1531 ATGGATGATTTAAAGAGGGATCTTTGTTACAATTTATTTGGTGGTACTAATAAACAATTT 1590
Qy 3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTTACTTTGTTGGGAGCCCTGTGACC 3188
Db 1591 ACAAAAGGTGGTGGTTATAGAGGTGATATAAATATTTATTTATGTTGGTATCATCTACT 1650
Qy 3189 AGTAAATCTCAGCTGCTTCAAGTATGTCACAGATAGCTCTCTGTCGGAATCTACACTAGT 3248
Db 1651 TCTAAATCAAAATCTTACAATATGTTTCAAGATTTGCTCCAAAGAGGGTTTATATCTTCT 1710
Qy 3249 GGGCGAGGAAGTTTCGGCGGTTCGAGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db 1711 GGTAAAGGTCTTCGGCTGTTGGTTGACAGCTTATATCACAGAGATATTGATACAAA 1770
Qy 3309 GAGACGGTATTCGAGAGCGGAGCTTTGGTTCTTTAGTGTGTCGGGATATGCTGTATGAT 3368
Db 1771 CAATTAGTATTGGAAGTGGGGCCCTTGTGTTATCTGATGAGGAGTGTGTTGTTGTTATGAT 1830
Qy 3369 GAGTTCGACAAAATGCTGATTAATGCCGAGCATCTTCAATGAGGTAAATGGAGCACAA 3428
Db 1831 GAGTTTGACAAGATGATGATGCCACCATCAGTTTTCATGAAGTCATGGGAACAA 1890
Qy 3429 ACGGTATCTGTAGCCAAAAGGGGTATCATTTGCCCTCGCTGAAACGCTCGGAGCTGTGCTCT 3488
Db 1891 ACTATTTCTATTGCTTAAAGCCGGTATTTATACCAATTAATGCTAGGACATCTATTTTAA 1950
Qy 3489 GCATGTGCAAACTCTTAGTGGGTCCCGATACAATGCGCGGCTTTTCTGTGATGTGAATCAATC 3548
Db 1951 GCTTCTGCAAACTCAATTAATTTCTGTTATGATCCAAATTTGCTGTCGCGGGTAAACATTT 2010
Qy 3549 CAGCTTCTCCCAACTCTACTTCTAGATTTGATTTAAATTTACTTAATGCTCGACACCA 3608
Db 2011 GATTTGCCCAACCATTTGTTATCACGCTTTTGGTTTATTTGATTTTGGGATAAAGTT 2070
Qy 3609 GACGAGCAAAACGATCGTCTGCTCGCAGGCACTCTCGTGGCTTTTACACTATGAAAACTAT 3668
```

```
Db      2071  GATGAATCAATCGATAGACAAATTGGCTAGACATTTGACTGATATGTATTTGGAGATGCT 2130
Qy      3669  GAAGTTTCAAAGCAGGACGCCCTTAGATCTAC-----AAACACTTACCGGTATATFACCC 3722
Db      2131  CCTGAAACTGTTAATGCCAATTCCGTTATTACCGTGGAAATTAATAACACTTTATATTCAA 2190
Qy      3723  TATGTCGTGCGCATGTACATCTTACATTAAGTATGTAAGCTGCTGAAGATTTGATTAAT 3782
Db      2191  TATGCTAAAGAGAAATTTCAATCCCGTGTACAGAAAGGTAAGAAATGAATTTAGTTAGA 2250
Qy      3783  GGTATGTTGAGATGCGCCAAAGGCAACTTTCCTGGAAGCAGTAGTAAAAAGGTGATAACA 3842
Db      2251  TCATAGCTTGAAATGAGAAATTTGGGGGAGATGCTAGATCATCAGAAAGAGAAATCACC 2310
Qy      3843  GCCACACCTCGCAACTCGAAAGTATGATTCGATCAGTGAAGCCCTAGCTCGAATGAGA 3902
Db      2311  GCTACTACAAGACAAATTGGAATCGATGATTAGATTGCTGGAAGCTCATGCTAAAAATGAGA 2370
Qy      3903  TTTTCTGAAAGTGGTAGAGAAAGTTGATGACGACGAGAGCTGTGGCCCTTTTAGACGTGCT 3962
Db      2371  TTATCTGAAAGAGTTGAAATTTGATCGATGTTAAAGAAAGCAGTAAAGATTAATTAATCGGCC 2430
Qy      3963  TTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGATGATCTTATCAGACT 4022
Db      2431  ATTAAGATTTGCTACTGATCCAGTTACTGCTAGTATCGATAGATGATGATGATTCAAACT 2490
Qy      4023  GGAGTGTGCGCCAGCAGAGCGTATTTCGTCGGGCCAACTTGTAGCTGCTCTGCGAGAGCTT 4082
Db      2491  GGTACTACGGCACACAAAGAGAGTACAAGAGATTGGTCAGTGAATTAATGAAATC 2550
Qy      4083  ATAGCAGTAATAAATT 4097
Db      2551  ATCGAAGAAATAAT 2565

RESULT 13
US-10-369-493-46198
; Sequence 46198, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46198
; LENGTH: 2802
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-46198

Query Match      10.8%; Score 470.8; DB 6; Length 2802;
Best Local Similarity 55.6%; Pred. No. 3.2e-136;
Matches 948; Conservative 0; Mismatches 747; Indels 9; Gaps 2;

Qy      2364  AAGCATATTGAGGCCAGACCTTCAATCTCAAGAGCATCGGTGCACATGGGTGAATCAAC 2423
Db      895  AAGTTTTATAAAGTAAGGCCCTTCAATGTAGTTCCTGTAAGGTATGCGCGAATTGAAT 954
Qy      2424  CCTTCAGATACAGCAAAATGGTTTCTGTTAAAGGAATGGTTATCGGGTCAGTTCTATC 2483
Db      955  CCAATGATATTGATTAATTTGATAAATTAAGGGTCTGCTCTGAGATCAATCCGGTT 1014
Qy      2484  ATACCTGAAATTAAGGGGCCCTTCTTCAATGTTTAGTGTGTGTCCTCGCTCGCCTA 2543
```

```
Db      1015  ATCCCTGATATGAAGGTAGCGTGTTCCTCAATGCAAGCTCTCGCATCA---TACAATGGCA 1071
Qy      2544  GTTACAGATTGTTAAAGGCGCGGTGAGGAGCCAAACAAGGTGTGAAGAAGCAGAAATGTGCA 2603
Db      1072  GTGGAATTTGATAGAGGAGTTATACAAGAGCCCGCTAGGTGTGAACGTATTGATTGTAAT 1131
Qy      2604  GCACGGATGCTATGCTCTTATTCACAAATCGATGCACATTTTGCAGAAATGAAGCAGATAGT 2663
Db      1132  GAACCCAAATTCATGTCTGATTCACAAAGGTGTTCAATTCGACATAAACAAGTCAAT 1191
Qy      2664  CGTCTTCAAGAAACTCCAGATGCCATTCCTGGAAGGAGAGACTCCACACACAGTCAGCATG 2723
Db      1192  AAGTTACAGAAACTCCAGACTTTTGCTGCTGATGGAACAAACGCTCACTCTATCTATTA 1251
Qy      2724  TGTTTATACAAACATATGTTGATGCTGTGAAGCTGGAGATGCTATGAGGTAAACAGGA 2783
Db      1252  TGTGTTTACGATGAATTAGTGGATTCTTGACGGCGGGCGATCGTATTGAGGTGACTGGC 1311
Qy      2784  GTTTTCAAGGCCATGGCAGTTGCTGAGTTGGTCCGAATCAACGAACATTAACGAGCATTTGAT 2843
Db      1312  ACGTTTCAGGTCCATCCCATTTAGAGCTAATTCAGGCAACGGGTACTTAAAGTCGTTGAT 1371
Qy      2844  AAGACCTACATCGATTGCGTGCACGTCAGAGAGTCTGACAGGGGTGCTACTGCAAACTGAA 2903
Db      1372  AAAACATACCTCGATGCTGCTCCAGTTAAAGTTTCAAGATTAAGAGTTAGACGTGCGAT 1431
Qy      2904  GATCCTATGAGATGGAATGAAGAGATGATATGTATGCTGGGTATCATGAAGTGTACT 2963
Db      1432  ACTTCTACTATTGAACAAGAAATTAATGCAAGAACAGGTAGATCATACGAGGTGCAAGAA 1491
Qy      2964  TCAGAAAGCTGCTAAATGAAGCAAGATTCAAAACCTTAAAGAGCTGTCCAAGTCCCGGGC 3023
Db      1492  GTAAGACAAATTAAGTCAAGGATTTAGCAAAATTTCCGAGAGGTTCCGCAAGAGAGAT 1551
Qy      3024  ATTTATGATAGACTTTCAAGGTGCGTGGCTCCAAGCATTTGGAGCTTGAAGATATTAAA 3083
Db      1552  TTATACAGTTTATTAGCCCGCTCTATTGCCCAAGTATTTACGACTAGAGATGTCAAG 1611
Qy      3084  AAGGGTCTTCTTTGCGCAGCTCTTTGGTGGGAAGGCTAAGAAAAATTCATCTCGAGCATCT 3143
Db      1612  AAGGGTATATTACTTTCAGCTATTGGCGCACGAATAAAACCTTTTACCAAGGTGGCCGT 1671
Qy      3144  TTCGAGGTGACATCAATGTTTAACTTTGTTGGGACCCCTGGTACCAAGTAAATCTCAGCTG 3203
Db      1672  TATAGAGGTGACATAAATATTTTACTTTGTGGGGATCTCTTACTTTCCAAATCGCAAT 1731
Qy      3204  CTTCAATGTCACAAAGATAGCTCTCGTGGAACTCTACACTAGTGGCGAGGAAGTTCG 3263
Db      1732  TTGCAATACGTTACAAAAATTTACTCTCGTGGTGTATATCTTCGGGTAAAGGTTCATCT 1791
Qy      3264  GCGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAAACTCGAGAGACGGTATTGGAG 3323
Db      1792  GCGTTGGTTTAACTGCTTATATTACAAGGATGTGCAACAAAAACAATTGTTTGGAA 1851
Qy      3324  AGCGGAGCTTTGTTGTTCTTAGTATCGTGGGATATGCTGTATCGATGAGTTCGCAAAATG 3383
Db      1852  AGTGTGTCATTAGTATTGCTGATGGAGGTGTTGTTGTTGTTGATTCGAGATTGATAAAATG 1911
Qy      3384  TCTGATAATGCCGAGCATGCTTCATGAGGTAATCGAGCAACAAACGGTATCTGTAGCC 3443
Db      1912  AGTGAATTTACAAAGGTCCGCTTTGCAAGGATCATGGAACAGCAGACTATTTCAATCGCA 1971
Qy      3444  AAAGGGGTATCAATTCGCTCGCTGAACGCTCGGACGCTGTCTCTGTCATGTCGCAAAATCT 3503
Db      1972  AAAGCGGAAATTAACAACACTTAATGCAGAAATTTCTATTTTGGCCAGTGTCAACCCA 2031
Qy      3504  AGTGGTCCCGATACAAATGCGGCCCTTCTGTGATTGATAAACAATCCAGCTTCTCCAACT 3563
Db      2032  ATTGTTTACGCTTACAAACCCCAATTTGCTGTGACTGAAAAATATTGATCTTACCGCCCCA 2091
Qy      3564  CTACTTCTAGATTTGATTTAATTTAATCTTAACTGTCGACAAACACGACGAGGCAAAAGAT 3623
Db      2092  CTACTTTCGAGATTTCGATCTGCTGCTATCTTGTCTTGTGATAAGGTTGATGAGAAAAATGAC 2151
```


QY 3624 CGTGGTCTCGCAGGACATCTCGTGGCTTTACATATGAAAA-----CTATGAAGTTTCA 3677
DB 2152 AGAGAACTAGCCAAACACTTAACAAATCTTTACCTGGGAAGATAAGCCCGAACATATTTCT 2211
QY 3678 AAGCAGGACGCTTAGATCTCAAAACACTTACCGGATATATCACCTATGCTCGTCAGCAT 3737
DB 2212 CAAGACGACGCTGCTACCAAGTTGAATTTTAAACGATGTATATTTAGTTATGCAAAAGGACAC 2271
QY 3738 GTACATCTACATTAAGTGAAGTCTGTAAGATTTGATTAATGGCTATGTTGAGATG 3797
DB 2272 ATACACCCAAATATCAGGAGCCCTAAGACTGAGCTGTGCTGCTTATGTAGGAATG 2331
QY 3798 CGCCAAAAGGGCAACTTTCTGGAAAGCAGTAAAGAGGTATAACAGCCACACCTCGGCCAA 3857
DB 2332 AGAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAAATCACAGCTACACAAAGACAA 2391
QY 3858 CTCGAAAGTATGATTCGATATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGTA 3917
DB 2392 CTTGAAAGTATGATTCGTTGGCTGAGGCGCACGCCCAAAATGAAATTTGAAAAACGTCGTA 2451
QY 3918 GAGAAGTTGATGACGACGAGAGCTGTGCGCTTTTAGACGTCGCTTTCGAGCAATCTGCT 3977
DB 2452 GAGCTGGAGGATGTTCAAGAGCCGTTAGATTAATTAGATCAGCCATAAAGATTTATGCA 2511
QY 3978 ACTGATCATGCAACAGGTACGATAGACATAGCATCTTATCACGACTGGAGTGTGCGGCCAGC 4037
DB 2512 ACAGACCTTAAACCGGTAAATTCACATGAATTTAGTTCAACAGGTAAATCAGTTATT 2571
QY 4038 GAGCGTATTCGTCGGGCCCACTTG 4061
DB 2572 CAGAGAAAACACTACAGGAGATTG 2595

RESULT 14

US-10-128-714-2558
; Sequence 2558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT FILING DATE: 2002-04-23
; PRIOR FILING DATE: 2002-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2558
; LENGTH: 2625
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-2558

Query Match 10.4%; Score 451.6; DB 5; Length 2625;
Best Local Similarity 56.1%; Pred. No. 3.4e-130;
Matches 944; Conservative 0; Mismatches 719; Indels 21; Gaps 4;
QY 2382 CCTTTCAATCTCAAGCATCGGTGCACATCGGTGAACCTCAACCCCTCAGATATAGACAAA 2441

DB 724 CCTTTGGTCTCGACTCGATGTGAAATATGAGACCTCGATCTCGACGATGGATAA 783
QY 2442 TTGTTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGG 2501
DB 784 CTAGTAAGCATTAAGGGCTTAGTCAATCGAAGCAGACCCATCAATCTCTGATATGAAGAA 843
QY 2502 GCCTTCTTCAAAATGTTTAGTGTGTGTCACCTCGCTCGCTAGTTTACAGTTGTTAAAGGG 2561
DB 844 GCCTTCTTCCGTTCCCAAGTCTGCAACCATGGTGTTCAGGTTCGACA-----TTGATCGTGA 900
QY 2562 CGGTTGAGGAGGCAACAAAGGTGTGAAAGCCAGATGTGACGACCGAAATGCTATGTCT 2621
DB 901 AAGATTGGGAACCCACTGAGTGCACGCTCAGTGTGTAAGGAACGAAATCTCGATGCAA 960
QY 2622 CTTATTCAATCATGATGACATCTTTTGCAAAATAGCAGATAGTGCCTCTTTCAAGAAATCTCA 2681
DB 961 CTATCCATAACCGCTGTGTTATTTGCCGACAAGCAGGTCAATCAAGTTGCAGGAACACCT 1020
QY 2682 GATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACTATG 2741
DB 1021 GACAGCATTCCTGATGCGCAGACTCTCACTCGGTTTCCTTTGTGTGTATGATGAGCTG 1080
QY 2742 GTTGATGCTGTGAAGCCCTGGAGATCGTATTGAGGTAAACAGAGTTTTCAGAGCCCATGGCA 2801
DB 1081 GTGATGTCTCAAGGGCTGGTGTGATCGGGTGAAGTGAACCGGTATTTTCGGTGCACACCT 1140
QY 2802 GTTCAGTTGCTCGGAATCAACGAACATTAACGAGCATTTGATTAAGACCTTACATCGATTGC 2861
DB 1141 GTGCGGTTAATCTCGCCAGGTTACACAGAAGTCGCTGTTCAAGACGTACATAGATGT 1200
QY 2862 GTGCACGTCAAGAGTCTGACAGGGGTGCACTGCAAACTGAAGATCCTATGAGATGAT 2921
DB 1201 CTTATGTTTCAGAAAGATCGATCGCAAGAAAGTTGGGTATCGAGCTCTCGACCATCGAGCAG 1260
QY 2922 AAGGAGATGATGATGATGCTGGGTATCATGAAGTGATCTTCAAGAGCTCTCTAATGAA 2981
DB 1261 GAGCTCTCGGAACAGCGCGTGGGATGCAGAAACAAACACAGTAGCTCTCTCGGAGGAG 1320
QY 2982 GCMAAGATTTCAAAAACCTTTAAAGAGCTGTCAGCTCCCGGGCATTTATGATAGACTTCA 3041
DB 1321 GAAGAG-----AAATTAAGCGAACTGCTACAGACCTGATCTGTATGAGCTTCTCTCT 1374
QY 3042 AGTCTGCTGGCTCAAGCATTTGGGAGCTTGAAGATATTAATAAGGCTTCTTTGCCAG 3101
DB 1375 CGGTCTTGGCCCCCAGCATCTACGAGATGGAACGACGTGAAGAGGGAATCTCTGCTTCAG 1434
QY 3102 CTCCTTTGGTGGGAAGGCTAAGAAATTCATCTGGAG-----CATCTTTCCGAGGTGAC 3155
DB 1435 TTGTTTGGAGGCACCAACAAAGACCTTTCAGAAAGGGTGGTAACCCAGATACCGTGGAGAT 1494
QY 3156 ATCAATGTTTTACTTTGTTGGGACCTCGTACACAGTAAATCTCAGCTGCTTCAGTATGTG 3215
DB 1495 ATCAATATCTTCTCTGTGGTGACCCATCTACATCAAGTCCAGCTTCTTCGTTACGTC 1554
QY 3216 CACAAGATAGTCTCTCGTGGAAATCTACACTAGTGGCGGAGGAAGTTTCGGGGTTGGCTG 3275
DB 1555 CATAGATTGCCCTCGCGGTGTATACAGCGGCAAGGGCTCTCTCGGCTGTGTGCTT 1614
QY 3276 ACAGGTATGTAACGAGGATCCAGAACTCCGAGAGACGTTATGGAGAGCGGAGCTTTG 3335
DB 1615 ACGGGTAGCTCACCGCGCATCTTGAAACCCGCGAGATGCTCTCGAGTCGGGTGCTTGG 1674
QY 3336 GTTCTTAGTGTGCTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCC 3395
DB 1675 GTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAAGTGAACGAATCCACT 1734
QY 3396 CGAAGCATCTTCATGAGGTAATGGAGCAACAAACCGGTATCTGTAGCCAAAGGGGTATC 3455
DB 1735 CGGTCCGTTCTGCATGAAGTCAATGGAACAAACACAGATATCTATCGCCAGGCGGCAAT 1794
QY 3456 ATTGCTCTGCTGAACGCTCGGAGCTGTGCTCTTTCATGTGCAATCTCTAGTGGTCTCCGA 3515

Db 1795 ATCACTACTTTGAACGCTAGACACGATCTCGCTTCGCGCAATCCGATCGGTAGCAGG 1854
Qy 3516 TACAATGCGCGCTTTCTGTGATTTAATCAATCCAGCTTCTCTCAACTCTCTACTTTCTAGA 3575
Db 1855 TACAATCCCAACTTGGCGGTTCTCTCAAAATATTGACCTTCCGCTACCTTGTCTCTCCGA 1914
Qy 3576 TTTGATTTAATTTACTTTAATGCTCGAACAACGAGACGAGAAAAGATGCTGCTCGCC 3635
Db 1915 TTCGACTTGGTATACCTCGTCTGGACCGAGTGGATGAGCAGGAAGATCGTGGCTCGCT 1974
Qy 3636 AGCAGCTCGTGGCTTTACACTATGAAAC-----TATCAAGTTTCAAGCAGGAGGCC 3689
Db 1975 AAGCACTTGTCAATATGACCTTGGAGACAGACCTGAGCATGCTGCCGAGCAAGAAATC 2034
Qy 3690 TTAGATCTCAAAACACTTACCGGTATATCACTATGCTCGTCAGCATGTATCATCTTACA 3749
Db 2035 TTGCGGATCGAATTCCTTACAGCTATATCACTACGCGCAAGACCAAGTCCATCCAGTG 2094
Qy 3750 TTAAGTATGAAGCTGTGAAGATTGATTAATGAGTATGTTGAGATGCGCCAAAAGGGC 3809
Db 2095 CTCACACCGCGCCGCTAAAGCTTTGTCGGATGCTTACCGTTAAACATGCGTAAGCTTGA 2154
Qy 3810 AACTTTCTTGAAGCAGTAAGAGTGATAAACAGCCACACCTCGGCAACTCGAAAGTATG 3869
Db 2155 GATGACATCGGTCTTCTGACCGCGTATCACCGCTACCACTCGTCAACTGGAGTCCATG 2214
Qy 3870 ATTCTGATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGAT 3929
Db 2215 ATCCGACTGCGAAGCGCATGCGGTATGCGGTATCGCGGAGGTCACTGCGGATGAT 2274
Qy 3930 GCAGCAGAGCTGTGCGCTTTTAGAGCTGCTTTGAGCAATCTGCTACTGATCATGCA 3989
Db 2275 GTGAGAGAACCGTGGCGCTGATCCGCTCGCGCATCAAGCAGGCGGCCACTGACTCTCGG 2334
Qy 3990 ACAGGTACGATAGACATGATCTTATCAGCACTGGAGTGTGCGCCAGCGAGCGTATTCGT 4049
Db 2335 ACCGGTCTGATGACATGAGCTTGTGACGAGGGGCACTAGTCCAGCAGACGCGAGC 2394
Qy 4050 CGGG 4053
Db 2395 CGGG 2398

RESULT 15

US-10-128-714-7558
; Sequence 7558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastian M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7558
; LENGTH: 3072

; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-7558
Query Match 10.4%; Score 451.6; DB 5; Length 3072;
Best Local Similarity 56.1%; Pred. No. 3.8e-130; Indels 21; Gaps 4;
Matches 944; Conservative 0; Mismatches 719;
Qy 2382 CTTTCAATCTCAAGCATCGGTGCAATCGGTGAACTCAACCTTCAGATATAGACAAA 2441
Db 1171 CTTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCTCGAGACATGGATAA 1230
Qy 2442 TTGGTTTCTGTTAAAGGAATGTTATCCGGTGCAGTCTTATCATATCATCTGAAATTAAGGG 2501
Db 1231 CTAGTAGAATTAAGGGCTTAGTCAATTCGAAACACACCCATCAATCTGTATATGAAGAA 1290
Qy 2502 GCCTTCTTCAAAATGTTTGTGTGTGTCTCACTCGCTCCGCTAGTATCAAGTTGTTAAAGGG 2561
Db 1291 GCTTTCTTTCGTTGCCAAGTCTGCAACCATGTTGTTGAGTGCACA---TTGATCGTGA 1347
Qy 2562 CGGTTTGAGAGCCAAACAAGGTGTGAAAGCCAGAAATGTGCAGACGGAATGCTATGTCT 2621
Db 1348 AAGATTGCGGAACCCACTGAGTCCCGACGTCAGTGTGTAAGGAACGAAACTCGATGCAA 1407
Qy 2622 CTTATTCATATCGATGCACTTTTGCATAATGACAGATAGTGTCTTTCAAGAAACTCCA 2681
Db 1408 CTCATCCATAACCGCTGTGTATTTTGGCGACAGAGTCAATCAAGTTGCGAGGAACACCT 1467
Qy 2682 GATGCCATTCCTGAAGGAGAGACTCCACACAGTCAAGTGTGTTTATACAACTATG 2741
Db 1468 GACAGATTCCTGATGCGCAGACTCTCTCTCGTTTCCGTTTGTGTGTATGATGAGCTG 1527
Qy 2742 GTTGATGCTGTGAAGCCTTGAGATCGTATTGAGGTAAACAGAGTTTTTCAAGGCCATGGCA 2801
Db 1528 GTGATGCTGCAAGGCTGTGTATCGGTGGAAGTGACCGGTATTTTCCGGTGCACCCCT 1587
Qy 2802 GTTCGATGTTGTCGGAATCAACGAACATTACGAGCATTTGTAAGACCTACATCGATTGC 2861
Db 1588 GTGCGGTTAAATCTCGCCAGCGGTACACAGAAAGTCTGCTGTTCAAGACGTACATAGATGTT 1647
Qy 2862 GTGCAGTCAAGAGTCTGACAGGGTTCGACTCCAACTCAAGATCCTATGGAGATGAT 2921
Db 1648 CTTTCATGTTCAAGAAATCGATCGCAAGAGTTGGGTATCGACGTCTCGAACATCGAGCAG 1707
Qy 2922 AAGGAGAATGATATGATGCTGGGTATCATGAAGTGATATCTTCAGAAAGTCTGCTAAAGAA 2981
Db 1708 GAGCTCTCGGAACAGCGGCTGGGATGAGAACAAACACAGTAGGCTCACTCGGAGAG 1767
Qy 2982 GCAAGATTCAAAAATTAAAGAGCTGTCCAAGTCCCGGGCATTTATGATAGACTTTCA 3041
Db 1768 GAAGAG-----AAAATTAAAGCGAACTGCTACAGACCTGATCTGTATGAGCTTCTCTCT 1821
Qy 3042 AGGTGCTGGTCCAGCATTTGGGAGCTTGAAGATATTAAGAGGTCTTCTTTGCGCAG 3101
Db 1822 CGGTCTCTGGCCCCCAGCATCTACGAGATGAGACGATGAAGAGGGAATCCCTGCTTCAG 1881
Qy 3102 CTCCTTTGGTGGGAAGGCTAAGAAAATTCATCTCGGAG-----CATCTTCCGAGGTGAC 3155
Db 1882 TTGTTTGGAGGACCAACAGACTTCCAGAGGGTGGTAACCCACATACCGTGGAGAT 1941
Qy 3156 ATCAATGTTTACTTGTGGGAGCCCTGGTACAGTAAATCTCAGCTGCTTCAGTATGTG 3215
Db 1942 ATCAATATCCTTCTCTGTGTGACCCCATCTACATCCAAAGTCCAGCTTCTTCGTTAGTCT 2001
Qy 3216 CACAAGATAGCTCTCTGTGGAATCTACATAGTGGGGAGAGAGTTTCGGCGGTGGGCTG 3275
Db 2002 CATAAGATTGCCCTCTCGCGTGTGTATACCGCGGCAAGGGCTCTCTCGGCTGTGGCTT 2061
Qy 3276 ACAGCGTATGTAAGAGGATCCAGAACTCCAGAGACGGTATTGGAGAGCGGAGCTTGTG 3335
Db 2062 ACGGCGTACCTACCCGCGATCTTGAAACCCGACAGATGTTCTTCGAGTCGGGTGCTTGT 2121
Qy 3336 GTTCTTAGTATCGTGGGATATGCTGTATCGATGAGTTTCGACAAATGTCTGTAATGCC 3395

[illegible]

Search completed: December 6, 2005, 09:52:34
Job time : 2140 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2005, 02:17:11 ; Search time 395 Seconds
(without alignments)
3425.784 Million cell updates/sec

Title: US-10-768-511-5
Perfect score: 4348
Sequence: 1 atggcgcgccactacgt.....gcgcaattccagagctgc 4348

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3289935 seqs, 155610033 residues
Total number of hits satisfying chosen parameters: 6579870

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA_New.*
1: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186.4	4.3	3769	7	US-11-087-227-89
2	180	4.1	2821	7	US-11-087-227-13
3	180	4.1	2900	7	US-11-087-227-15
4	173.2	4.0	2715	6	US-10-507-275-4
5	160.8	3.7	3453	7	US-11-087-227-11
6	157.6	3.6	3379	6	US-10-507-275-2
7	123.2	2.8	507	7	US-11-087-172-158
8	58.2	1.3	1851	6	US-10-750-185-45476
9	54.8	1.3	1312	6	US-10-750-185-55669
10	39.6	0.9	876	6	US-10-750-185-27656
11	37.6	0.9	449	6	US-10-821-234-393
12	35.6	0.8	978	6	US-10-821-234-409
13	35.4	0.8	3150	6	US-10-793-626-4034
14	35.4	0.8	4231	6	US-10-793-626-4115
15	34.2	0.8	1263	6	US-10-793-626-3317
16	34.2	0.8	1410	6	US-10-750-185-632786
17	34.2	0.8	2247	6	US-10-750-185-48673
18	34.2	0.8	4116	6	US-10-793-626-4100
19	34	0.8	1210	6	US-10-793-626-41443
20	33.4	0.8	2957	6	US-10-793-626-4135
21	33	0.8	1758	6	US-10-750-185-38232
22	33	0.8	2428	6	US-10-750-185-50035
23	32.8	0.8	1003	6	US-10-750-185-37272

ALIGNMENTS

RESULT 1

US-11-087-227-89
; Sequence 89, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89
; LENGTH: 3769
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-087-227-89

Query Match 4.3%; Score 186.4; DB 7; Length 3769;
Best Local Similarity 51.7%; Pred. No. 3.2e-50;
Matches 477; Conservative 0; Mismatches 436; Indels 9; Gaps 2;

QY 3025 TTTATGATGACITTTCAAGTCCTGGCTCCAGCAATTTGGGAGCTTTGAAGATATTAATAA 3084
Db 1113 TATACCACATCTTTGTACCAGCCTGTTCCCTACTATACATGCAATGATGAAGTAAAC 1172
QY 3085 AGGGCTCTTTTCCAGCTCTTTGGTGGGAAGCGCTAAGAAATTCATCTGGAGCATCTT 3144
Db 1173 GGGGTGCTCTGCTGATGCTCTTTGGTGGCGTTTCCAAAGACACAGGAGAGGGACCTCTC 1232
QY 3145 TCCGAGGTGACATCAATGTTTACTTTGGGAGCCTGCTACAGTAAATCTTCAGCTGC 3204
Db 1233 TTCAGGGGACATAAATGTTTGCATTGTTGTGACCCCAAGTACAGTAAAGCCATTTTC 1292
QY 3205 TTCAGTATGTCACAGATAGCTCTCGTGGAAATCTACACTAGTGGCGGAGGAAGTTCCG 3264
Db 1293 TCAGCACGTGGAGGAGTTTCAGCCCAAGAGCTGTCTACACAGTGGTAAAGCGTCCAGTG 1352
QY 3265 CGGTGGCTGACAGCGGTATGTAAAGAGATCCAGAAATCTCAGAGACGGTATTGGAGA 3324

Db 1353 CTGCTGGCTTAACAGCAGCTGTTGTGAGAGATGAAGATCTCATGATTTGTTCATTGAGG 1412
 Qy 3325 GCGAGCTTTGTTCTTAGTAGTACGTGGGATATGCTGATCATGATGAGTTGCGACAAATGT 3384
 Db 1413 CTGAGCTTTGATGTTGGCTGATAAATGGTGTGTTGTTATGATGAATTTGATAAGATGG 1472
 Qy 3385 CTGATAATGCCGAGCATGCTTCATGAGTAATGGAGCAACAAACGGTATCTGTAGCCA 3444
 Db 1473 ACGTGGGGATCAAGTTGCTATTTCATGAAGCTATGGAAACAGCAGACCATATCCATCACTA 1532
 Qy 3445 AAGGGGGTATCATTTGCTCGCTGAACGCTCGGACGCTCTCTCTGCTGATGCAAAATCCTA 3504
 Db 1533 AAGCAGAGTGAAGCTACTCTGAAGCCCGGAGCTCCATTTTGGCAGCAGCAACCCCA 1592
 Qy 3505 GTGGGTCCGATACAAATGCGCGCTTTCTGTGATGATGAACATCCAGCTTCCTCAACATC 3564
 Db 1593 TCAGTGGACACTATGACAGATCAAAATCATTTGAACAGAAATATAAAATTTGTCACTCCCA 1652
 Qy 3565 TACTTTCTAGATTTGATTTAAATTTACTTAATGCTGACAAACAGCAGCAAAACGATC 3624
 Db 1653 TCATGTCCGATTCGATCTCTTTTATCTTTGTGGATGAATGTAAATGAGGTTACAGATT 1712
 Qy 3625 GTGCTCTCCGAGCATCTCGGCTTTTACACTATGAAATATGAAGTTTCAAAGCAGG 3684
 Db 1713 ATGCCAATGCCAGGCCATAGTAGATTGC---ATTCAAGAAATTGAGGAATCAATTGATC 1769
 Qy 3685 AGCCCTTAGATCTACAAACACTTTACCGCGTATATCACTATGCTCGTCAAGCATGTACATC 3744
 Db 1770 GTGCTATTCCCTCGATGATATCAGAAGATATCTTCTTTGCAAGACAGATTAAAC--- 1826
 Qy 3745 CTACATTAAGTAGTGAAGCTGCTGAAGATTGATTAATGGCTATGTTGAGATCGGCCAA 3804
 Db 1827 ---CCAAGATTTCCAAGAGTTCAGAGGACTTCAATTTGTGGAGCAATATAAACATCTCCGCC 1883
 Qy 3805 AGGGCAACTTTCCTGGAAGCAGTAAAGGTGATAACAGCCACCTCGGCACTCGAAA 3864
 Db 1884 AGAGAGATGTTCTGAGTGACCAAGTCTTCATGGAGGATTACAGTGCGACAGCTTGAGA 1943
 Qy 3865 GTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAAG 3924
 Db 1944 GCATGATTCGTCTCTGAAGCTATGGCTCGGATGCACTGCTGTGATGAGGTTCCAACCTA 2003
 Qy 3925 TTGATGACGAGAGCTGTGCG 3946
 Db 2004 AACATGTGAAGAGCTTTCCG 2025

RESULT 2

US-11-087-227-13
 ; Sequence 13, Application US/11087227
 ; Publication No. US20050260566A1
 ; GENERAL INFORMATION:

; APPLICANT: Fischer, Timothy J.
 ; APPLICANT: Malinowski, Douglas P.
 ; APPLICANT: Taylor, Adrian J.
 ; APPLICANT: Parker, Margaret R.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
 ; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
 ; FILE REFERENCE: 046143/287139
 ; CURRENT APPLICATION NUMBER: US/11/087,227
 ; CURRENT FILING DATE: 2005-03-23
 ; PRIOR APPLICATION NUMBER: 60/556,495
 ; PRIOR FILING DATE: 2004-03-24
 ; NUMBER OF SEQ ID NOS: 90
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 2821
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens

US-11-087-227-13

Query Match 4.1%; Score 180; DB 7; Length 2821;
 Best Local Similarity 51.2%; Pred. No. 3.3e-48;

Matches 479; Conservative 0; Mismatches 445; Indels 12; Gaps 2;
 Qy 3020 GGGCAATTTATATAGACTTTCAAGGTGCTCGCTCCAAGCATTTTGGGAGCTTGAAGATAT 3079
 Db 1500 GGATTTCTACGAAAGCTGGCAGCTTCAATCGCCCCAGAAATATATACGGGCATGAAGATGT 1559
 Qy 3080 TAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGCTAAGAAAAATTCATCTCTGGAGC 3139
 Db 1560 GAAGAGGCACTGTGCTCTCTGCTAGTCGGGGGTGG---ACCACTCTCTCGAGGCAT 1616
 Qy 3140 ATCTTTCCGAGGTGACATCAATGTTTACTTTGTTGGGACCTGTGTACAGTAATCTCA 3199
 Db 1617 GAAATCCGGGGCAACATCAACATCTGTCTGATGGGGATCCTGCTGTGGCCCAAGTCTCA 1676
 Qy 3200 GCTGCTTCAAGTATGTGCAAGATAGCTCTCTGCTGGAATCTACATAGTGGGGGAGGAG 3259
 Db 1677 GCTCTCTCATACATTGATCGACTGTGCGCTCTGCGACCGAGTACACAACAGGCGCGGGCTC 1736
 Qy 3260 TTGCGGGGTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGCTATT 3319
 Db 1737 CTCAGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTCCGCTGAGTGGAGAACTGACCTT 1796
 Qy 3320 GGAGAGCGGAGCTTTTGGTTCTTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAA 3379
 Db 1797 AGAGGGTGGGGCTTGGTGGCTGACAGGGTGTGCTGCAATTGATGAGTTCGACAA 1856
 Qy 3380 AATGCTGANTAAATGCCAGAGCATGCTTCATGAGTAAATGGAGCAACAAACGGTATCTGT 3439
 Db 1857 GATGCTGAGGCGGACCGCACAGCCATCCACGAGGTCATGGAGCAGCAGACCATCTCCAT 1916
 Qy 3440 AGCCAAAGGGGTATCATTTGCTCGTGAACGCTCGGAGCTGTCTTGTGATGTCGAAA 3499
 Db 1917 TGCCAAAGGCGGCACTTCCACCACTCAATGCCGCTGCTCATCTGCTGCTGCCCAA 1976
 Qy 3500 TCTTAGTGGGTCCCGATACAATGCGCGCTTTCTGTGATGATAACATCAGCTTCTCTCC 3559
 Db 1977 CCTGCTACGGGCGCTACACCCCTCGCGCAGCCTGGAGCAGAAACATACAGTACTCTGC 2036
 Qy 3560 AACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTGCAACAAACAGCAGCAAAA 3619
 Db 2037 TGCACTGTCTCTCCGCTTTGACCTCTCTGGCTGATTCAGGACCGGCCCGACCGAGCAA 2096
 Qy 3620 CGATGCTGCTCGCCAGGCATCTCGTGGCTTTTACACTATGAAAACATATGAAGTTTCAAA 3679
 Db 2097 TGACCTTACGGTTGGCCCGACATACCTATGTGACCCAGCAGCAGCGGCGAGCCCCCTC 2156
 Qy 3680 GCAGGACGCTTAGATCTACAAACACTTACCGGTATATCACCTATGCTCGTCAGCATGT 3739
 Db 2157 CCAGTTTGAACCTCTGGACATGAAGCTCATGAGGCGTTACATAGCCATGTGCCGAGAA 2216
 Qy 3740 ACATCTCATATTAAAGTGAAGCTGCTGGAAGATTTGATTAATGGCTATGTTGAGATGCG 3799
 Db 2217 GCAGCCCATGGTCCAGAGTCTCTGGCTGACTACATCAGCAGCAGCATACGTTGAGATGAG 2276
 Qy 3800 CAAAAGGGCAACTTTCCTGGAAGCAGTAAAGGTGATAACAGCCACACCTCGGCAACT 3859
 Db 2277 GCGAGAGG-----CTTGGGCTAGTAAGGATGCCCTATACCTTCTGCCCGAGCCCT 2327
 Qy 3860 CGAAAGTATGATTCGTTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGTGAGA 3919
 Db 2328 GCTGGCTATCTCGGCTTTCCACTGCTCTGGCACGCTTGAGAAATGTTGATGTTGGTGA 2387
 Qy 3920 GAAAGTTGATGACGAGAGAGCTGTGCGCTTTTAGA 3955
 Db 2388 GAAAGAAAGTGTGAATGAAGCCATCAGGCTTAATGGA 2423

RESULT 3

US-11-087-227-15
 ; Sequence 15, Application US/11087227
 ; Publication No. US20050260566A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fischer, Timothy J.

APPLICANT: Malinowski, Douglas P.
APPLICANT: Taylor, Margaret R.
APPLICANT: Parker, Margaret J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
DETECTION OF CERVICAL DISEASE
FILE REFERENCE: 046143/287139
CURRENT APPLICATION NUMBER: US/11/087,227
CURRENT FILING DATE: 2005-03-23
PRIOR APPLICATION NUMBER: 60/556,495
PRIOR FILING DATE: 2004-03-24
NUMBER OF SEQ ID NOS: 90
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 2900
TYPE: DNA
ORGANISM: Homo sapiens
US-11-087-227-15

Query Match 4.1%; Score 180; DB 7; Length 2900;
Best Local Similarity 51.2%; Pred. No. 3.3e-48;
Matches 479; Conservative 0; Mismatches 445; Indels 12; Gaps 2;
Qy 3020 GGGCAATTATGATAGACTTCAAGTCTGCTCGCTCCAGCAATTTGGGAGCTTGAAGATAT 3079
Db 1579 GGATTTCTACGAAGCTGGCAGCTTCAATCGCCCAAGAAATATACGGGCATGAAGATGT 1638
Qy 3080 TAAAAAGGCTCTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAATTCATCTGGAGC 3139
Db 1639 GAAGAAGCACTGTCTCTCTGCTAGTCGGGGGTGTGG--ACCACTCTCTCGAGGCAT 1695
Qy 3140 ATCTTTCGAGGTGACATCAATGTTTACTTGTGGGGACCTGGTACCAATTAATCTCA 3199
Db 1696 GAAATCCGGGGCAACATCAACATCTGTCTGATGGGGATCTGTGTGGCCAAAGTCTCA 1755
Qy 3200 GCTGCTTCAGTATGTGCACAGATAGCTCTCTCTGGAATCTACATAGTGGCGGAGGAG 3259
Db 1756 GCTCTGTCAATGATGCTAGCTGGCGCTTCGAGCCAGTACACAGGCCGGGGCTC 1815
Qy 3260 TTCCGGGCTGGGCTGCAGCGTATGTAAAGGAAGATCCAGAACTCGAGAGACGGTATT 3319
Db 1816 CTCAGGAGTGGGCTTACGGCAGCTGTGCTGAGAGACTCCGTGAGTGGAGAACTGACCTT 1875
Qy 3320 GGAGAGCGGAGCTTTGGTCTTCTAGTATCGTGGGATATGCTGTATCGATGAGTTGCAAA 3379
Db 1876 AGAGGGTGGGCGCTTGGTGTGCTGACCGGGTGTGTGCTGCTATGATGAGTTGCAAA 1935
Qy 3380 AATCTGTATATGCCCGAAGCATGCTTCAAGGATGATGAGCAACAAACGGTATCTGT 3439
Db 1936 GATGGCTGAGCCGACCGCAAGCCATCCAGAGGTCAAGGAGCAGCAGCAATCTCCAT 1995
Qy 3440 AGCCAAAGGGGATATCATTTGCTCGCTGAACGCTCGGAGCTGTGCTTGCATGTGCAAA 3499
Db 1996 TGCCNAGCGCGCATTTCCACCACTCAATGCGCGTGTCCATCTCTGGTGGCGCA 2055
Qy 3500 TCCTAGTGGGTCCCGATCAATAGCGCGCTTTCTGTGATGATGATAACATCCAGCTTCCTCC 3559
Db 2056 CCCTGCTTACGGGCGCTACAACTCGCGCGAGCTCGGAGCAGAACATACAGCTACCTGC 2115
Qy 3560 AACTCTACTTCTAGATTTGATTTAAATTAATGCTGCAAAACAGACGAGCAAAA 3619
Db 2116 TGCACTGCTCTCCGGTTTGAACCTCTCTGCTGATTCAGGACCGGCGCCGACGAGCAA 2175
Qy 3620 CGATCTGCTCTCGCAGGATCTCTGGCTTTTACACTATGAAACTATGAATTTCAAA 3679
Db 2176 TGACCTACGGTTGGCCAGCATACCTATGTGCAACGACAGCGCGGAGCCCCCTC 2235
Qy 3680 GCAGAGCGCTTAGATCTTACAAACACTTACCGCGTATACCTATGCTCGTCAGCATGT 3739
Db 2236 CCAGTTTGAACTCTGGACATGAAGCTCATGAGGGTTACATAGCCATGTGCCCGGAA 2295
Qy 3740 ACATCTTACATTAAGTATGATGCTGTGAAGATTTGATTAATGCTATGTTGATGATCG 3799
Db 2296 GCAGCCCATGTTGCCAGAGTCTCTGGCTGACTATACATCAAGCAGCATACGTGGAGATGAG 2355

Qy 3800 CCAAAGGGCAACTTTCTCGAAGCAGTAAAGAGGTGATTAACAGCCACACTCGCAACT 3859
Db 2356 GCGAGAGG-----CTTGGGCTAGTAAGGATGCCACTATATCTTCTGCCGAGCCCT 2406
Qy 3860 CGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGA 3919
Db 2407 GCTGGCTATCTCGCCCTTTCACATGCTCTGGCAGCTGAGAAATGGTGGATGTGGTGA 2466
Qy 3920 GAAAGTTGATGCAGCAGAAAGCTGTGCGCCTTTTGA 3955
Db 2467 GAAAGAGATGTGAATGAAGCCATCAGGCTAATGGA 2502

RESULT 4

US-10-507-275-4

; Sequence 4, Application US/10507275

; Publication No. US20050250166A1

; GENERAL INFORMATION:

; APPLICANT: Masai, Hisao

; APPLICANT: Tamai, Katsuyuki

; APPLICANT: Medical and Biological Laboratories Co., Ltd.

; APPLICANT: Japan Science and Technology Agency

; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.

; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods

; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK

; TITLE OF INVENTION: Kinase Inhibitory Ability

; FILE REFERENCE: 082368-001100US

; CURRENT APPLICATION NUMBER: US/10/507,275

; CURRENT FILING DATE: 2004-09-09

; PRIOR APPLICATION NUMBER: JP 2002-067702

; PRIOR FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918

; PRIOR FILING DATE: 2003-03-12

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: Patentin ver. 2.1

; SEQ ID NO 4

; LENGTH: 2715

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(2715)

US-10-507-275-4

Query Match 4.0%; Score 173.2; DB 6; Length 2715;

Best Local Similarity 53.4%; Pred. No. 5.5e-46;

Matches 364; Conservative 0; Mismatches 318; Indels 0; Gaps 0;

Qy 3006 CTGTCCAAAGCTCCCGGCGATTATGATAGACTTTTCAAGTCTGGCTCCCAAGCAATTTGG 3065
Db 1399 CTCTCCAAGATCAGCAATTTGGAGAGAGATCTTGGCAGCANTTGCACCTCCATCTAT 1458
Qy 3066 GAGCTTGAAGATATTAAGAGGGTCTTTCTTTCAGAGCTCTTTTGGTGGGAAGGCTAAGAAA 3125
Db 1459 GGGCATGAAGACATCAAGAGAGGCTGCTGCTCTGCGCTTGTGGAGGGGAGCCCAAGAAC 1518
Qy 3126 ATTCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTACTTGTGGGGAGCCCTGGT 3185
Db 1519 CCAGGTGGAAGCAACAAGTTTCGAGGTGACATTAATGTGCTTGTGTGGGGAGCCCTGGC 1578
Qy 3186 ACCAGTAAATCTCAGCTCTTCACTATGTGCACAGATAGCTCTCGTGGATCTACACT 3245
Db 1579 ACAGCAAGTCCCAATTTCTCAATATCATCAGAAAGTGTAGCCGTGCCATCTTCACC 1638
Qy 3246 AGTGGCGAGGAATTCGGCGGTTGGGCTGACAGCTGTAAAGAGATCCAGAAACT 3305
Db 1639 ACTGCCAGGTGGTGGTCTCAGCAGTGGGTCTCACCAGGTACGTTCAGCGGCATCCCGTCAGC 1698
Qy 3306 CGAGAGACGGTATTTGGAGAGCGGAGCTTTGGTCTTATGATGATCGTGGGATGCTGTATC 3365
Db 1699 AGAGAGTGGACCTTTAGAGGCGGAGCGCTGGTCTTGGCTGTACCGGGGGGTGTGTCTCATT 1758

QY	3366	GATGAGTTTCGACAAATGTCTGATTAATGCCGAAGCATGCTTTCATGAGTAAATCGAGCAA	3425
Db	1759	GACGAGTTTGACAAGATGAATGACCCAGGACAGGACAGCATCCAGAGGCCCATGGAACAG	1818
QY	3426	CAAACGGTATCTGTACCCAAAGGGGGTATCATTTGGCTCGCTGGAACGCTCGGACGCTCTGTC	3485
Db	1819	CAAAGCATCTCCATCTCCAAAGCGTGGCATGTTCCTCGCTGCAAGCCCGCTGCACCTGTC	1878
QY	3486	CTTGCATGTGCAAAATCCTATGTGGGTCCCGATACAATGGCGCGCTTTCTGTGATTGATAAC	3545
Db	1879	ATAGTCTGCTGCCAAACCCCATAGGAGCGCTACGACCTTCACTGACCTTCTCAGAGAAAT	1938
QY	3546	ATCCAGCTTCTCCAACTCTACTTTCTAGATTGTTTGAATTTACTTAATGCTCGACAAA	3605
Db	1939	GTAGACCTCACAGAGCCCATCATTTCCCGCTTTGATGTCTCTGTGTGGTGAGGGACACT	1998
QY	3606	CCAGACGAGCAAAACGATCGTCTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAAAC	3665
Db	1999	GTTGATCCAGTTCAGGATGAGATGCTGGCCCCGCTTTGTGTTGGCAGCCACGTCAGACAC	2058
QY	3666	TATGAAGTTTCAAGACGAGCAGC	3687
Db	2059	CACCCAGTAAACAGAAGGATG	2080
RESULT 5			
US-11-087-227-11			
; Sequence 11, Application US/11087227			
; Publication No. US20050260566A1			
; GENERAL INFORMATION:			
; APPLICANT: Fischer, Timothy J.			
; APPLICANT: Malinowski, Douglas P.			
; APPLICANT: Taylor, Adriann J.			
; APPLICANT: Parker, Margaret R.			
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE			
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE			
; FILE REFERENCE: 046143/287139			
; CURRENT APPLICATION NUMBER: US/11/087,227			
; CURRENT FILING DATE: 2005-03-23			
; PRIOR APPLICATION NUMBER: 60/556,495			
; PRIOR FILING DATE: 2004-03-24			
; NUMBER OF SEQ ID NOS: 90			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 11			
; LENGTH: 3453			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-11-087-227-11			

Query Match	3.7%;	Score 160.8;	DB 7;	Length 3453;
Best Local Similarity	50.9%;	Pred. No. 7.8e-42;		
Matches 381;	Conservative 0;	Mismatches 367;	Indels 0;	Gaps 0;
QY	2940	GCTGGGTATCATGAAGTGATACATCTTCAGAGCTGCTAATGAAGCAAGATTCACAAAACCTT	2999	
DB	1390	GCCAAAGAAGCAACAAGGTTCCTCTAGCGGGAACTGACCCGATGAAGATGTGAAGATGATC	1449	
QY	3000	AAAGAGCTCTCAAGCTCCCGGGCATTTATGATAGACTTTTCAAGGTGCGCTGGCTTCCAAGC	3059	
DB	1450	ACTAGGCTCTCAAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATGTGCTCTCTCC	1509	
QY	3060	ATTTTGGGAGCTTGAAGATATTAAAAAGGCTCTTCTTTGCCAGCTCTTTTGGTGGGAAGGCT	3119	
DB	1510	ATCTATGGTCATGAAGACATCAAGAGAGCCTGGCTCTGGCCCTGTTCCGAGGGGAGGCC	1569	
QY	3120	AAGAAAATTCATCTCGAGCATCTTTCCAGGTGACATCAATGTTTTACTTGTTTGGGGAC	3179	
DB	1570	AAAAACCCAGGTGGCAAGCACAAAGGTACGTGGTGATATCAACGTCTCTTCTTGCGCGAGAC	1629	
QY	3180	CCTGGTACCMGTAAATCTCAGCTGCTTCCTAGTATGTGCACAAGATAGCTCCCTCGTGGAACT	3239	
DB	1630	CCTGGGCACAGCAAGTCGCAAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCGGAGCCATC	1689	

3240	QY	TACACTAGTGGCGAGGAAGTTCCGGCTTGGCTGACAGCGTATCTAAAGGAGATCCA	3299
1690	Db	TTCAACACTGGCCAGGGGCGCTGGCTGTGGGCGCTCACGGCGTATGTCMAGCGCACCTT	1749
3300	QY	GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTTGGTTCTTAGTGATCGTGGGATATGC	3359
1750	Db	GTCAAGCAGGAGTGGACCTTTGGAGGCTTGGGGCCCTGGTTCTGCTGACCGAGGAGTGTGT	1809
3360	QY	TGTATCGATGAGTTGCACAAAATGTCGTGATAATGCCCGAAGCATGCTTCATGAGGTAATG	3419
1810	Db	CTCATTGATGAATTTGCACAAGATGAATGACCAGGACAGAACCCAGCATCCATGAGGCCCATG	1869
3420	QY	GAGCACAAACGGTACTGTGAGCCAAAGGGGTATCATTCGCTCGCTGGAACGGCTCGGACG	3479
1870	Db	GAGCAACAGAGCATCTCCATCTCGAAGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGC	1929
3480	QY	TCGTCTCTTCGATGTGCAATCTCTAGTGGGTCCGATACAAATGGCGGCCCTTTCGTGATTT	3539
1930	Db	ACGGTCATTGCTGCGGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCT	1989
3540	QY	GATAACATCAGCTTCTCCAACTCTACTTTCTAGATTGCAATTAATTTACTTAATGCTC	3599
1990	Db	GAGAACGTGGACCTTCACAGAGCCCATCATCTCAGCTTTGACATCCCTGTGTGTGGTAGG	2049
3600	QY	GACAAACACAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACATAT	3659
2050	Db	GACACCGTGACCAGTCCAGGACGAGATGCTGSCCCGCTTCGTGGTGGGCGAGCCACGTC	2109
3660	QY	GAAACTATGAAAGTTTCAAAGCAGGACG	3687
2110	Db	AGACACCAACCCAGCAACCAAGGAGGAG	2137

```

RESULT 6
US-10-507-275-2
; Sequence 2, Application US/10507275
; Publication No. US2005020166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhibitory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JPO03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31)..(2709)
US-10-507-275-2

```

	Query Match	Best Local Similarity	3.6%;	Score 157.6;	DB 6;	Length 3379;
	Matches 379;	Conservative	50.7%;	Fred. No. 8.7e-43;	Indels 0;	Gaps 0;
Qy	2940	GCTGGGTATCATGAAGTGATCTTCAGAGCTGTTATGAGCAAGATTCAAAACCTT	2999			
Db	1330	GCCAAGAGGACACAAAGGTGTGCTGTAGGGAACTGACCGCATGAAGATGTGCAAGTATGC	1389			

QY 3000 AAAGAGCTGTCAAGCTCCCGGCAATTTATGATAGACTTTCAAGTTCGCTCGCTCCAGC 3059
Db 1390 ACTAGCTCTCCAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATGCTCTCTCC 1449
QY 3060 ATTTGGGAGCTTGAAGATATAAAGAGGTCTTCTTGGCCAGCTCTTTGGTGGGAGGCT 3119
Db 1450 ATCTATGTCATGAAGACATCAAGAGAGGCCCTGCTCTGGCCCTGTTCCGAGGGAGGCC 1509
QY 3120 AAGAAAATTCATCTGGAGCATCTTTCCAGGTGACATCAATGTTTACTTGTGGGAC 3179
Db 1510 AAAAACCAGGTGGCAGCACAAGGTACGTGTGTATATCAAGTCTCTTTGGCGAGAC 1569
QY 3180 CTTGTACAGTAATCTCAGCTCTTCAATATGTCAGATATGTCAGAGTCTCTCTGTAATC 3239
Db 1570 CTTGGCAGCAGCAAGTCGAGTCTTCTCAAGTATATTTGAGAAAGTGTCCAGCCGAGCCATC 1629
QY 3240 TACACTAGTGGCGAGGAGTTCGGCGGTGGGCTGACAGCGTATGTAACGAGATCCA 3299
Db 1630 TTCACACTGGCCAGGGGCGTCCGCTGTGGCCGTACGGCGTATGTCCAGCGGACCCCT 1689
QY 3300 GAAACTCGAGAGACCGTATTTGAGAGCGGAGCTTTGGTCTTGTAGTATCGTGGGATATGC 3359
Db 1690 GTCAGCAGGAGTGAACCTTGGAGGCTGGGCGCTGTCTTGGCTGACCGAGGAGTGT 1749
QY 3360 TGTATGATGATTCGACAAAATGTCGATAATGCCGAGCATGCTTTCATGAGGTAATG 3419
Db 1750 CTCATGATGATTTGACAAAGATGAATACAGGACAGAACCCAGCATCCATGAGGCCATG 1809
QY 3420 GAGCAACAAACGGTATCTGTACCAAGAGGGGTATCATGCTCGCTGAACGCTCGAGC 3479
Db 1810 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACCTCCCTCGAGGCTCGTC 1869
QY 3480 TCTGTCTTGCATGTGCAAACTCTAGTGGGTCCGATACAAATGCGCGCTTCTCTGTAT 3539
Db 1870 ACGTCACTGTGCGCCCAACCCCATAGAGGGGCGCTAGGACCCCTCGTGAATCTTCTCT 1929
QY 3540 GATAACATCCAGCTTCTCCAACTCTACTTTCTAGATTTGATTTAAATTTACTTAATGCTC 3599
Db 1930 GAGAACGTGGACCTCAGAGAGCCCATCTCTCAGCTTTGACATCTCTGTGTGTGAGG 1989
QY 3600 GACAAACAGAGCAGCAAAACGATCGTCTGCGCAGGAGATCTCGTGGCTTTTACATAT 3659
Db 1990 GACACCGTGGACCCAGTCAGAGCAGAGATGCTGGCCCGCTTCTGTGTGGGAGCCAGCTC 2049
QY 3660 GAAACTATGAATTTCAAAAGCAGGACG 3687
Db 2050 AGACACACCCAGCAGCAGGAGG 2077

RESULT 7

US-11-108-172-158

; Sequence 158, Application US/11108172
; Publication No. US20050260177A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon B.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedrick Thomas S.
; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.471C15
; CURRENT APPLICATION NUMBER: US/11/108,172
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 10/025,380
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 09/922,217
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/833,263
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 09/649,811
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/609,448
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/575,251
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/519,444
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 09/504,629
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US 09/480,321
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: US 09/476,296
; PRIOR FILING DATE: 1999-12-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 27, 428, 448, 449, 456, 462, 490, 492, 497, 498, 502, 503
; OTHER INFORMATION: n = A,T,C or G

US-11-108-172-158

Query Match

2.8%; Score 123.2; DB 7; Length 507;

Best Local Similarity 61.8%; Pred. No. 4.4e-30;

Matches 207; Conservative 0; Mismatches 122; Indels 6; Gaps 1;

QY 2992 AAAAATTAAGAGCTGTCCAGCTCCCGGCAATTTATGATAGACTTTCAAGTTCGCTCG 3051
Db 167 AATTGCTTTAAGAACTTTCCAGGAAACAGACATTTATGAGAGGCTTCTTCAGGCTTGG 226
QY 3052 CTCCAAGCATTTGGGAGCTTCAAGATATTTAAAGGGTCTTCTTTGCCAGCTCTTTGGTG 3111
Db 227 CTCGAGCATTTATGAACATGAAGATATAAGAGGGAATTTTGCTTCAGCTCTTTGGCG 286
QY 3112 GGAAGGCTTAAGAAAATT-----CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTT 3165
Db 287 GGACAAGGAAGGATTTTAGTCACACTGGAAGGGCAAAATTTCCGGCTGAGATCAACATCT 346
QY 3166 TACTTGTGGGAGCCCTGGTACCAAGTAATCTCAGCTGCTTCAGTATGTGCACAAGATAG 3225
Db 347 TGCTGTGTGGCGACCTGGTACCAAGTCCAGCTGCTGCAGTACGTGTACAACTCG 406
QY 3226 CTCCTCGTGGAACTCTACACTAGTGGCGGAGGAAGTTCCGCGGTTGGGCTGCAGCGTATG 3285
Db 407 TCCCAGGGGCCAGTACACATGTTGGAGGAGGCTCCAGTGCANNTGGCCTNACTGCTNACG 466
QY 3286 TAACGAAGGATCCAGAAACTCGAGAGCGGTATTG 3320
Db 467 TAATGAAGACCTTGAGCAAGGNACTGGNNCTG 501

RESULT 8

US-10-750-185-45476

; Sequence 45476, Application US/10750185

; Publication No. US20050260603A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard

Thu Dec 8 10:36:56 2005

```

; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45476
; LENGTH: 1851
; TYPE: DNA
; ORGANISM: Bovine 19866880565574
US-10-750-185-45476

Query Match      1.3%; Score 58.2; DB 6; Length 1851;
Best Local Similarity 59.3%; Pred. No. 2.8e-08;
Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 3285 GTAACGAGGATCCAGAACTCGAGACGGTATTTGGAGCGGAGCTTTGTTCTTAGT 3344
Db 1381 GTAACGCTGTAAAGACTCAGAGAAATTTGGAGCGGAGCATTAGTCTCTCGCA 1440

Qy 3345 GATCGTGGGATATCTGTATGATGATGTTCCACAAAATGCTGATAATGCCCGGAAGCATG 3404
Db 1441 GATCGTGGCTCTGCTGTATGATGATGTTTAAAGAGCATGACAGAACTAGT 1500

Qy 3405 CTTCAATGAGTAATGAGCAACAAACGGTATCTGTAGCCAAAGGGG 3451
Db 1501 ATCCAGAGCAATGAGCAGCAACCACTAGTGTGTCTAAGGCTGG 1547

RESULT 9
US-10-750-185-55669
; Sequence 55669, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55669
; LENGTH: 1312
; TYPE: DNA
; ORGANISM: Bovine 19866880618533
US-10-750-185-55669

Query Match      1.3%; Score 54.8; DB 6; Length 1312;
Best Local Similarity 61.0%; Pred. No. 2.8e-07;
Matches 89; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 3238 TCTACACTAGTGGCGAGGAAGTTTCGGCGGTGGGTGACGCGCTGTACGAGCATGTACGAGGATC 3297
Db 841 TGTACAGTCTGGGAAAGGAGCAGCGCGCGGCTGACAGCCCTCGGTGATGAGGGACC 900

Qy 3298 CAGAACTCGAGACGGTATTTGGAGCGGAGCTTTGTTCTTAGTGATCGTGGGATAT 3357
Db 901 CTTGCTCCGGAACCTTCATCATGAGGCGGGGCCATGTTCTTGCCGACGAGGTGTGCG 960

```

```

Qy 3358 GCTGTATCATGATGAGTTCCACAAAATG 3383
Db 961 TCTGTATTGATGAGTTTCACAAAGTG 986

RESULT 10
US-10-750-185-27656
; Sequence 27656, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27656
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Bovine 19866881490050
US-10-750-185-27656

Query Match      0.9%; Score 39.6; DB 6; Length 876;
Best Local Similarity 58.5%; Pred. No. 0.021;
Matches 69; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1260 CTTCAAGAACTCCAGATGCCATTCCTGAAGGAGACTCCACACAGTCAGCATGTGT 1319
Db 508 CTCAGGAGTCCCTCTGAAGACATGCCCGCGGGCAGACGCCGACACAGTGTCTCTTT 567

Qy 1320 TTATACACACTATGTTGATGCTGTGAAGCTCGAGATCGTATTGAGGTAAACAGGAG 1377
Db 568 GCTCACAATGACCTCGTGGACAAAGTCCAGCTGGGACAGAGTCACGTACAGGTG 625

RESULT 11
US-10-821-234-393
; Sequence 393, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 393
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(449)
; OTHER INFORMATION: n = a,t,c or g
US-10-821-234-393

Query Match      0.9%; Score 37.6; DB 6; Length 449;
Best Local Similarity 54.3%; Pred. No. 0.058;

```


Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 4115 TGCTTTGAAGACAGTCAGCTCTTCTGAGGATATCCGGAGCCAAAGCAGTGTGACGTTAG 4174

Db 65 TGTCTCCAGAGCTGTCCAGGAATTTGGTCTTGCCCGGTTCAAAAGCAAGCTGACCAAGAC 124

QY 4175 TTTCAGGATATTAATAATGCTCTGGTAGCCTCCAAAGGAGAGGCTTTCTTACTGTCCA 4234

Db 125 TATGAAGGGTTTGAATATATCTTGGCTAAGCTGCAAGGCGAGGCCCTTCCAAACACT 184

QY 4235 TGTGACATAGTCAAGAGAG 4254

Db 185 TGTTGAGACGCCAAGGGAG 204

RESULT 12

US-10-821-234-409/c

; Sequence 409, Application US/10821234

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: Labat, Ivan

; APPLICANT: Stache-Grain, Birgit

; APPLICANT: Andarmani, Susan

; APPLICANT: Tang, Y. Tom

; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia

; FILE REFERENCE: 821A

; CURRENT APPLICATION NUMBER: US/10/821,234

; CURRENT FILING DATE: 2004-04-07

; PRIOR APPLICATION NUMBER: US 60/462,047

; PRIOR FILING DATE: 2003-04-07

; NUMBER OF SEQ ID NOS: 1704

; SOFTWARE: pt_seq_genes Version 1.0

; SEQ ID NO 409

; LENGTH: 978

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-821-234-409

Query Match 0.8%; Score 35.6; DB 6; Length 978;

Best Local Similarity 55.7%; Pred. No. 0.47;

Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 2994 AAATTAAAGAGCTGCCAGCTCCCGGCATTTATGATAGACTTTCAAGTTCGCTGGCT 3053

Db 519 ATACTGTTGAGCTCTCCACTCTCATGCGCCAGCGATGATCACTCCGCCAGGGTCCCCA 460

QY 3054 CCAAGCATTTGGAGCTTGAAGATATTAAGAGGCTCTTTGCCAGCTCTTTGGTGGG 3113

Db 459 AACAGCACTGCTGATTTTAGAGTCATTGCAAGGGATCTTCATGTAGGGCTCAATTGTGC 400

QY 3114 AA 3115

Db 399 AA 398

RESULT 13

US-10-793-626-4034/c

; Sequence 4034, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4034

; LENGTH: 3150

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-4034

Query Match 0.8%; Score 35.4; DB 6; Length 3150;

Best Local Similarity 51.6%; Pred. No. 1.3;

Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 2886 GGTCCAGTGCACAACTGAAGATCCTATGGAGATGGATAGGAGAAATGATATGTATGCTGGG 2945

Db 2236 GGTAGAAATATGAAGATCATGCTAGTGAATTTAAATCATGAAGTTGATAGTTTATACGTA 2177

QY 2946 TATCATGAAAGTGATACTTCAAGAGCTGCTATGAGCAAGAAAGTTCAAAAACCTTAAGAG 3005

Db 2176 TTTACTAAAGCGGCTTCATCTTTAACTGGTGAAGCAACAATTCCTTAATCATAAAGAT 2117

QY 3006 CTGTCCAAGCTCCCGGCATTTATGATAGACTTTTCAA 3042

Db 2116 ATTACAGAACAAATTGGACTATGAGCGGAACCTTGGAA 2080

RESULT 14

US-10-793-626-4115

; Sequence 4115, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4115

; LENGTH: 4231

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-4115

Query Match 0.8%; Score 35.4; DB 6; Length 4231;

Best Local Similarity 51.6%; Pred. No. 1.6;

Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 2886 GGTCCAGTGCACAACTGAAGATCCTATGGAGATGGATAGGAGAAATGATATGTATGCTGGG 2945

Db 1768 GGTAGAAATATGAAGATCATGCTAGTGAATTTAAATCATGAAGTTGATAGTTTATACGTA 1827

QY 2946 TATCATGAAAGTGATACTTCAAGAGCTGCTATGAGCAAGAAAGTTCAAAAACCTTAAGAG 3005

Db 1828 TTTACTAAAGCGGCTTCATCTTTAACTGGTGAAGCAACAATTCCTTAATCATAAAGAT 1887

QY 3006 CTGTCCAAGCTCCCGGCATTTATGATAGACTTTCAA 3042

Db 1888 ATTACAGAACAAATTGGACTATGAGCGGAACCTTGGAA 1924

RESULT 15

US-10-793-626-3317

; Sequence 3317, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2005, 01:29:25 ; Search time 10575 Seconds
(without alignments)
19236.902 Million cell updates/sec

Title: US-10-768-511-5
Perfect score: 4348
Sequence: 1 atggcgccgcactacgt.....gcgcaattccagagctcgc 4348

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	871.2	20.0	2699	4	CNS0A83A
2	777	17.9	3168	10	CL959068
3	478	11.0	879	8	DR495517
4	469.6	10.8	881	7	CO123277
5	456.4	10.5	1156	7	CK210237
6	439.4	10.1	907	8	DR503163
7	434.2	10.0	770	2	BF051102
8	433	10.0	788	3	BM412106
9	429.2	9.9	2522	10	AY408033
10	428	9.8	775	3	BM410852
11	408.4	9.4	2357	10	AY408034
12	404.2	9.3	798	6	CP449139
13	394.4	9.1	709	8	DR917805
14	392.8	9.0	3285	4	AK089999
15	392.8	9.0	3301	4	AK088796
16	392.8	9.0	3545	4	AK036653
17	392.6	9.0	2519	10	AY408035
18	391.2	9.0	3316	4	AK011743
19	387.4	8.9	750	6	CA258367
20	379.6	8.7	706	3	BJ464173
21	378.6	8.7	765	6	CP442356
22	377	8.7	1082	5	BX836107

23	375.2	8.6	727	7	CK072837
24	374.8	8.6	755	8	DN232047
25	364.2	8.4	615	5	BQ465527
26	363.2	8.4	698	7	CV461490
27	358.4	8.2	717	6	CA501849
28	357.8	8.2	1607	4	CR598188
29	357.8	8.2	1833	4	CR606514
30	356	8.2	910	8	CK099871
31	352.4	8.1	619	3	BQ481103
32	349.2	8.0	638	5	BU989891
33	348.8	8.0	668	6	CA258647
34	342	7.9	644	7	CV008987
35	339.8	7.8	584	3	BQ481191
36	339.8	7.8	673	7	CV055224
37	336.6	7.7	642	6	CD475195
38	330.8	7.6	560	6	CD480657
39	329.8	7.6	823	8	CV881487
40	329	7.6	593	5	BQ459696
41	328.6	7.6	725	6	CA258377
42	327.8	7.5	632	8	CV794540
43	326	7.5	676	3	BP955291
44	322.6	7.4	571	3	BQ048021
45	319.2	7.3	613	6	CA827375

ALIGNMENTS

RESULT 1
CNS0A83A
LOCUS
DEFINITION
Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSLTPGH582F05 of Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress).
ACCESSION
BX820670.1 GI:42467030
VERSION
HTC; GSLT_cDNA.
KEYWORDS
Arabidopsis thaliana (thale cress)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 2699)
AUTHORS
Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C., Menard,M., Cruaud,C., Quetier,P., Scarpelli,C., Schachter,V., Temple,G., Cruaud,C., Weissenbach,J. and Salanoubat,M., Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 2699)
AUTHORS
Genoscope.
JOURNAL
Direct Submission
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich information center for Protein Sequences) . 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full length
http://www.genoscope.cns.fr/cgi-bin/ggb?ggb?source=Arabidopsis.
location/Qualifiers
1. .2699
/organism="Arabidopsis thaliana"
/mol_type="mrna"

FEATURES
source

/db_xref="taxon:3702"		/clone="GSLTPGH58ZF05"		/tissue_type="Hormone Treated Callus"		
/ecotype="Col-0"		/plasmid="pCMVSPORT 6"		complement(1..2699)		
/gene="At2g16440"						
ORIGIN						
Query Match		20.0%; Score 871.2; DB 4; Length 2699;				
Best Local Similarity		64.5%; Pred. No. 3e-257;				
Matches 1346; Conservative		0; Mismatches 718; Indels 24; Gaps 2;				
Qy	2178	AAGTACATCCAGATC	ATAGAGGAGACTGTG	GAGCGTGGAGGAGTACTCT	AAATATCGAC 2237	
Db	478	AAGTATATGGTTTC	GTATGAAGAAGGTG	ATGAGATTTGAAGAGATGG	ATGATGTTGAT 537	
Qy	2238	ATGTCAGACATTTAT	GACCATGATCTCT	GATCTATACGCAAAAT	TGTTGATACCCACTC 2297	
Db	538	CGCTTTGATGTTG	TGATATGATCCT	GTATTTGTATAACAGATG	TTGTTGTTATCCTTTG 597	
Qy	2298	GACATCATCCCCCT	GTTGGACACTGAG	TGTGAGGAAGTTGCTACT	CTCTTTACTACCAACG 2357	
Db	598	GAAGTTCTCGTAT	CTTCGATATTG	CTGCTGATGGATATT	GTGTTCCACGATTAATCGTTTG 657	
Qy	2358	TTTGAGAGCATATT	TGAGGCCAGACCTT	CAATCTCAAGCATCG	TGCACATCGTGAA 2417	
Db	658	TTGCGAAGAACAT	GTTCAAGTTAGGAT	TTTTTAACCTCAGACAT	CTACTTCCATGGAAT 717	
Qy	2418	CTCAACCTTCAGAT	ATAGACAAATTTGG	TTTCTGTTAAAGGAAT	GGTTATCCGGTGCAGT 2477	
Db	718	CTCAATCCATCT	GAATCGAGAAGAT	GTCTCTTTGAAAGGT	TATGATAATTCGGAGTAGC 777	
Qy	2478	TCTATCATACCT	GAAATTAAGGGG	CCCTTCTTCAAAAT	GTTTAGTGTGCTACCTCGCT 2537	
Db	778	TCTATCATCT	GAAATAGGGAAG	CTGTGTTAGATGCT	TGTTGTTGGTTACTTCTCT 837	
Qy	2538	CCGCTAGTTACAG	TGTTTAAAGGCGGG	TTGAGGACCAACAG	GTGTGAAAGCCAGAA 2597	
Db	838	GACCCCATCAT	CGTTGATAGGGA	AAATAAGTGAGCCT	CCTACTTGTGTTGAAACAGAG 897	
Qy	2598	TGTGCAGCAGGA	ATGCTATGCTCT	TTATTC	CAATTCGATGCACCTTTTGCAAAATAGCAG 2657	
Db	898	TGATGACCAAGA	ACTCCATGACACT	AGTGTGACACACCG	ATGCGATGCTGATAAGCAG 957	
Qy	2658	ATAGTCGTCTT	CAAGAACTCCAG	ATGCCATTCCTG	TGAAGGAGAGACTCCACACAGTC 2717	
Db	958	ATTGTGAGGCTT	TCAGGAACACCT	GATGAGATTC	GAGATTCGAGAGGAGAACACTCACACAGTT 1017	
Qy	2718	AGCATGTGTTTAT	ACAACTATGTT	GTGATGCTGTG	AAAGCCTCGAGATCGTATTGAGGTA 2777	
Db	1018	AGCTTGTGCT	ATGATAGT	AGTTGATTAAT	TGGGAAGCCTGGTATAGAAATGAGGTC 1077	
Qy	2778	ACAGAGTTTTT	CAAGGCCATGG	CAGTTGCGAGTT	GGTCCGAATCAACGAACATTTACGAGCA 2837	
Db	1078	ACCGGAATTT	TACAGAGCAAT	GACTGTCCGAGT	TGGGACCTGCTCACAGGACTGTGAAATCT 1137	
Qy	2838	TTGTATAGACT	CATCATGTTG	CGTGCACGT	CAAGAGTCTGACAGGGGTCCACTGCA 2897	
Db	1138	GTGTTTAA	GACCTATATCG	ATTGCTTCC	GAATCAACGAAGTAAGTTAAAGATGTCT 1197	
Qy	2898	ACTGAAGAT	CTCTATGG	AGATGGATA	GATATGTTATGCTGGGTATCATGAAAGT 2957	
Db	1198	GCTGAGGAT	CTTATG	ATGTTG	CAATAGTCTGCGTAGAGTTGATGAAGATGTCGAGTTA 1257	
Qy	2958	GATATCT	CAGAAGCTG	CTAATGA	AGCAAGATTC	AAAGAGCTGTCCAAAGCTC 3017
Db	1258	G-----	-----	-----	-----ATGAGAGAGAGCTGAGGAAGTTTCAGGAACCTCTTAACAA 1299	
Qy	3018	CCGGGCATTT	TATGATAG	ACTTTCA	AGGTTCGCTCC	AAAGCATTTTGGGAGCTTGAAGAT 3077
Db	1300	CCAGATATAT	ATGAGGCTT	AGTAGATCA	CTAGCACCAAC	ATCTGGGAGTTGGATGAT 1359

```
Db 2440 GGCAACATAAACACCGAAATTCATCTTCATGATGTTAGAAAAGCAGTCGCGACGCTAGCG 2499
Qy 4212 GGAGAAGGCTTTCTTACTGTCTCATCGTGTGACATAGTCAAGAGAGGTTTGA 4259
Db 2500 AGTGAAGGATTTCTGTGAGTCTGAAGGTGATAGAAATCAAGAGAGTATAA 2547

RESULT 2
LOCUS CL959068 3168 bp DNA linear GSS 21-SEP-2004
DEFINITION OaIFCC002182 Oryza sativa Express Library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.
ACCESSION CL959068
VERSION CL959068.1 GI:52372893
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 3168)
Ma L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
1..3168
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN
Query Match 17.9%; Score 777; DB 10; Length 3168;
Best Local Similarity 66.1%; Pred. No. 4.9e-228;
Matches 1178; Conservative 0; Mismatches 560; Indels 45; Gaps 2;
Qy 2217 GAGGATACCTAAATATGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCA 2276
Db 1282 GGGGAGTCGCTCGACGTCAACGCCACGACGTGTCGACACGACCCGACCTCTACGGC 1341
Qy 2277 AAAATGTTTCGATACCCATCGACATCATPCCCTCTGTTGGACATGAGTGTGAGGAGTT 2336
Db 1342 AAGATGGTCAGGTACCGCTCGAGGTGCTCGCATCTTTGATATCGTGTCTATGACCTC 1401
Qy 2337 GCTACTCTTTACTACCAAGTTTGAGAAGCATATGAGGCCAGACCTTTCAATCTCAA 2396
Db 1402 GTCCGCGCATCGAGCCCTCTTCGAGAAGCATCTCAGACAGGATCTACAACTCAAG 1461
Qy 2397 GCATCGGTGCATCGGTGAATCTCAACCTTCAGATATAGACAAATGGTTTCTGTAAA 2456
Db 1462 TCCTCGTTTGTGTGAGGAATCTCAACCTTCTGATATCGAAGATGGTGTCCATCAAG 1521
Qy 2457 GGAATGGTTATCCGGTGCAGTGTCTATCATACCTTGAAATTAAGGGGCCCTTTCTTCAAATG 2516
Db 1522 GGTATGATTAATTCGGTGCAGCTCGGTGATTCAGAGCTCAAGAGGCTGTGTTCCGTGC 1581
Qy 2517 TTATGTGTGGTCACTCGCCTCGGTAGTTACAGTTGTTAAAGGGGGGTTGAGAGCCA 2576
Db 1582 TTGGTTTGGGGTCTTACTCTGAGCCTGTAATGGTTGATCGAGGGGGGTTACTGAGCCT 1641
```

```
Qy 2577 ACAAGGTGTGAAAAGCCAGAAATGTGCAGACGGAATGCTATGTCTCTTATTACAAATCGA 2636
Db 1642 CACATCTGTGAGAAAGAAACAATGTAAAGCTACAAAATCTATGACTCTTGTGTAATACCGA 1701
Qy 2637 TGCATTTTTCGAAATAAGCAGATAGTGGCTCTTCAAGAAATCCAGATGCCATTCCTGAA 2696
Db 1702 TCCAGGTTTTCAGATAGCAGATCATAAAGTTGCAAGAAACACCAAGATGAGATACAGAA 1761
Qy 2697 GGAGAGACTCCACACACACAGTCAGCATGTGTTTATACAACACATATATGGTTGATGCTGTGAAG 2756
Db 1762 GGTGGCACTCCACATACCTGTTAGTGTCTTGATGACAGATAAGCTGTAGATCTCTGAAAG 1821
Qy 2757 CCTGGAGATCGTATTGAGGTAAACAGAGTTTTCAGAGCCATGCGAGTTCGAGTTGGTCCG 2816
Db 1822 CCTGGAGACAGGGTTGAGATAACTGGGATATACAGAGCCATGAGTATTAGAGTTGGCCCA 1881
Qy 2817 AATCAACGAACATACAGAGCATTTGTATGAACCTACATCGATTGCGTGCACGCTCAAGAG 2876
Db 1882 ACTCAGAGGACAGTGAAGTCGATTT-----TCAAGAAG 1914
Qy 2877 TCTGACAGGGTTCGACTGCAAACTGAAATCCTATGGAGATGATAAGGAGATATGATG 2936
Db 1915 ACAGACAAGTCTAGACTTCATGTTGAGGACTCCATGGAAAATGATAACCCCAATGCTA-- 1972
Qy 2937 TATGCTGGGTATCATGAAAGTGATATCTCAGAGCTGCTTAATGAAGCAAAAGATTCAAAA 2996
Db 1973 -----ACAAGACAACCTGAAGATGATTTTCTCAGAGATAAGGTTGAGAA 2016
Qy 2997 CTTAAAGAGCTGTCCAAGCTCCGGGCATTTATGATAGACTTTCAAGTCGCTGCTGCCA 3056
Db 2017 TTAAAGAGTTGTCAAGTTGCGAGATATATGACAGATTAACATAGGTGCTTCTGCTCCA 2076
Qy 3057 AGCATTTGGGAGCTTGAAGATATTAAGAGGCTCTTTTGGCCAGCTCTTTGGTGGGAAG 3116
Db 2077 AACATATGGAGCTGGAACGACGTTAAAGGGGCTCTTCTGCGAGCTTTTGGTGAAT 2136
Qy 3117 GCTAAGAAAATTCATCTGGAGCATCTTTCCGAGTGACATCAATGTTTTTACTTTGGG 3176
Db 2137 GCTTTGAGGCTTCTTCTGGAGTAGTTTCCGAGGCGACATCAATATATTTGCTTGTGT 2196
Qy 3177 GACCTGTGACAGTAATCTCAGCTGCTTCAGTATGTGCACAGATAGCTCTCTGTGA 3236
Db 2197 GATCCTGGAACAAGTAATCCAGCTTCTCCAATACATGCACAAACTGTCTCTCTGTGC 2256
Qy 3237 ATCTACACTAGTGGCGGAGGAAGTTCCGGCTGTGGCTGCACAGCGTATGTAACGAGGAT 3296
Db 2257 ATTTATACAGTGGCAGAGGAAGTTACAGCTGTTGGCTTACTGCAATACGTTACCAAGAT 2316
Qy 3297 CCAGAAACTCCGAGAGACGGTATTTGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGATA 3356
Db 2317 CCTGAAACTGGTGAACCTGTTCTTGAGAGTGGAGCGCTGTTTGTGATGACAAAGGTGT 2376
Qy 3357 TGCTGTATCGATGAGTTCGACAAATGTCTGATAATATGCCCGAAGCATGCTTCATGAGGTA 3416
Db 2377 TGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGTTACATGAGGTG 2436
Qy 3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCATTTGCTCTCGTGAACGCTCGG 3476
Db 2437 ATGGAAACACAGACTGTCTCCATTCGCAAGGCTGGAATAATTTGCAATCTTTAATGCCAG 2496
Qy 3477 AGCTCTGTCTTGCATGTGCAAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTG 3536
Db 2497 ACATCAGTCTTAGCATGTGCAAAATCCTACTGAATCAGTTAATCCAAAGGCTCTCTGTG 2556
Qy 3537 ATTGATAACATCCAGCTTCTCCAACTCTACTCTTCTAGATTTGATTTAATTTACTTAATG 3596
Db 2557 ATTGACAATATCCATCTTCTCCAACTCTCTTAGGTTTGAACCTCATTTATCTGATA 2616
Qy 3597 CTCGCAACACGACGAGCAAAACGATCGTCTGCGCAGGCACTCTCGTGGCTTTACAC 3656
Db 2617 TTGCAAGGACAGATGAGCAAACTGATAGACCCCTGGCTAAGCATATTTGTTTCTGTTGCAT 2676
```

QY 3657 TATGAAACATATGAAGTTTCAAGCAGGAGCGCTTAGATCTACAAACACTTACCGCGTAT 3716
 |||||
 Db 2677 TTTGAGAAATCCAAACATAGAGGAGCTGAGGTCTTGATTTGCCAACACTTGTAGCTAC 2736
 |||||
 QY 3717 ATCACTATGCTGCTGAGCATGTATCATCTTACATTAAGTGAAGAGCTGCTGAAGATTG 3776
 |||||
 Db 2737 ATAAATGATGCAAGAGCATATATCAACACAGTTATCTGATGAAGCTGCAGAAAGATTG 2796
 |||||
 QY 3777 ATTAATGCTATGTTGAGATGCGCAAGAGGCAACTTCTCTGGAAGCAAGTAAAGAGTG 3836
 |||||
 Db 2797 ACCCGTGCTATGTTGAGATGAGGAAGAGGAACAGCCCTGTGTAGCAGAAAGAGGTC 2856
 |||||
 QY 3837 ATACAGCCACACTCGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGA 3896
 |||||
 Db 2857 ATAACTGCGAGCTCGCAAAATGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCGA 2916
 |||||
 QY 3897 ATGAGATTTTCTGAAGTGGTAGAAGATTTGATGACGAGAGCTGTGCGCTTTTAGAC 3956
 |||||
 Db 2917 ATGCGATTTCTGAAGTGGTTGAAGTACGAGATGTTGTAGAGGCTTCAGGCTTCTCGAA 2976
 |||||
 QY 3957 GTCGGTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGA 3999
 |||||
 Db 2977 GTTGCCATGCGCAATCGGCACTGATCATGCCACTGCTTGA 3019
 |||||

RESULT 3

DR495517/c 879 bp mRNA linear EST 06-JUL-2005
 LOCUS WS0298_BR_E02 SS-IB-A-FL-15 Picea sitchensis cDNA clone WS0298_E02
 DEFINITION 3. mRNA sequence.

ACCESSION DR495517.1 GI:69453834

VERSION DR495517

KEYWORDS EST.

SOURCE Picea sitchensis (Sitka spruce)

ORGANISM Picea sitchensis

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

AUTHORS

Ralph, S., Koloosova, N., Oddy, C., Cooper, D., Butterfield, Y.,
 Kirkpatrick, R., Liu, J., Palmquist, D., Stott, J., Barber, S., Yang, G.,
 Babakaiff, R., Brown-John, M., Chand, S., Featherstone, R., Masson, A.,
 Mayo, M., Moran, J., Olson, T., Wong, D., Friedmann, M. P., Ritland, C. E.,
 Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B. E., Douglas, C.,
 Ritland, K., and Bohlmann, J.
 The spruce transcriptome: Analysis of expressed sequence tags from
 multiple cDNA libraries
 Unpublished (2005)

JOURNAL

Contact: Joerg Bohlmann
 Genome BC forest genomics program
 University of British Columbia
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
 Vancouver, British Columbia, Canada, V6T 1Z3
 Tel: 1-604-822-0282

Fax: 1-604-822-2114

Email: bohlmann@mel.ubc.ca

Plate: WS0298 row: E column: 02

High quality sequence stop: 879.

FEATURES

Source

1. .879
 /organism="Picea sitchensis"
 /mol_type="mRNA"
 /cultivar="FB3-425"
 /db_xref="taxon:3332"
 /clone="WS0298_E02"
 /sex="Hexamphrodite"
 /lab_host="E. coli DH10B cells"
 /clone_lib="SS-IB-A-FL-15"
 /note="Organ: Bark (with phloem and cambium attached) from
 two year old clonal trees grown under greenhouse
 conditions in standard potting soil mixture; Vector:
 pBluescript II SK (+) XR; Site 1: XhoI (5' end of cDNA);
 Site 2: BamHI (3' end of cDNA); Adult white pine weevils
 (Pissodes strobi) were caged with mesh bags on sapling

trees. Bark tissue with phloem and cambium attached was
 harvested 2 hours, 6 hours and 48 hours after continuous
 feeding by the insects. mRNA was isolated from each tissue
 source independently and equal quantities of mRNA from 20
 each tissue were then pooled. cDNA was prepared from 20
 micrograms of mRNA according to the full-length cDNA
 library construction method described by Carninci P. et
 al. (2000). Genome Research 10(10):1617-1630 and
 directionally ligated into the pBluescript II SK (+) XR
 vector digested with XhoI (5' end) and BamHI (3'). Plasmid
 DNA was then transformed by electroporation into DH10B
 cells (Invitrogen) for propagation."

ORIGIN

Query Match 11.0%; Score 478; DB 8; Length 879;
 Best Local Similarity 71.5%; Pred. No. 1.3e-135;
 Matches 628; Conservative 0; Mismatches 250; Indels 0; Gaps 0;
 QY 2415 GAATCAACCTTCAGATATAGACAAATGGTTCTGTTAAAGGAATGGTTATCGGTGC 2474
 |||||
 Db 878 GAATCAATCCATCAGATATAGACAAATGGTTCTGTTAAAGGAATGGTTATCGGTGC 819
 |||||
 QY 2475 AGTTCTATCATACCTGAAATTAAGGGGCGCTTCTTCAAATGTTTAGTGTGGTCACTCG 2534
 |||||
 Db 818 AGCTCAATAATACCTGAAATCAAAGAGGCATTTTTCAGATGCGCTTGTGTTGGCTATGCT 759
 |||||
 QY 2535 CCTCGGTAGTTACAGTTGTTTAAAGGGCGGTGTGAGGAGCCCAAGGTGTGAAAAGCA 2594
 |||||
 Db 758 CCACACAGAGCTGCTGTTGATAGAGTGAATTCATGAACTTCTCGATGTGGAAGGCCA 699
 |||||
 QY 2595 GAATGTGCAGCAGCAAGAAATGCTATGCTTTATTCATAATCATGATGCACTTTTGGCAATTAAG 2654
 |||||
 Db 698 GAATGTGCTGCTATGAACTCCATGACTTAGTTTATATATAGATGAGGTGTGCTGACAAA 639
 |||||
 QY 2655 CAGATAGTGCCTTCTCAAGAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
 |||||
 Db 638 CAGATCGTAAAGGCTGCAAGNAACCCCTGATGCTATTCAGAGGGAGAGAAACCCACATACA 579
 |||||
 QY 2715 GTCAGCATGTGTTTATACAACTATGCTGTGATGCTGTGAAGCCCTGGAGATCGTATTGAG 2774
 |||||
 Db 578 GTCAGCTTGTAAATGTCATGACAACTAGTTGATGCTGCCAAGCTGTTGACAGAGTTGAG 519
 |||||
 QY 2775 GTAAACAGGAGTTTTCAGGCCAATGCGATGCGAGTTGGTCCGAATCAAGCAACATTAACA 2834
 |||||
 Db 518 GTAAACAGGAATCTTCAGGGCCATGATGCTGTAGAGTAGGACCAACACAGAGGACAGTCAA 459
 |||||
 QY 2835 GCAITGTATAGACCTACATCGATTCGTCGACGCTCAAGAGCTCTGACAGGGGTGCACTG 2894
 |||||
 Db 458 TCCCTCTTTAAGACCTACATTGACTGCGCTTCACTTGAAGAGGCGAGACAAGACGACTG 399
 |||||
 QY 2895 CAAACTGAAGATCCTTATGAGATGCAATGAAGAGAAATGATATGATGCTGGGTATCATGAA 2954
 |||||
 Db 398 CAACTGAAGATCCAAATGAGGCGTGAATGCCACTAGTGACAGCTTGTGCTTCTTCAAGAG 339
 |||||
 QY 2955 AGTGATCTTCAAGAGCTGCTAATGAAGCAAAAGATTCAAAACCTTAAAGAGCTGTCCAAG 3014
 |||||
 Db 338 GGTGATATTTTCAACCGGTTTCAGTATGAAGCCAAAGATTGAGAAGTTGAAAGAGCTGTCA 279
 |||||
 QY 3015 CTCCTGGGCAATTTATGATAGACTTTCAAGGTGCTGCTGCCAAGCAATTTGGGAGCTGAA 3074
 |||||
 Db 278 CAGCCTGTATATATGAGAAGTTAAACAGGTCAATTCGGCTCCAAAGTATTTGGGAGCTG 219
 |||||
 QY 3075 GATATTTAAAAAGGCTCTTCTTTGCCAGCTCTTTTGGTGGGAAGGCTTAAGAAAAATTCATCT 3134
 |||||
 Db 218 GATGTGAAAAAGGGCTTTTATGTCAGCTTTTGGGGGAACCTCCAGTAAAGCTAGCCACA 159
 |||||
 QY 3135 GGAGCATCTTTCCGAGGTGACATCAATGTTTACTTTGTTGGGGACCTGTGTACAGTAAA 3194
 |||||
 Db 158 GGTGCCACCTTCGTTGGTGCATCAACATTTTCTGTTGTTGGGGACCCGGGTACGAGTAAG 99
 |||||
 QY 3195 TCTCAGCTGCTTCAGTATGTGCAAGATAGCTCCTCGTGGAAATCTACACTAGTGGGCGGA 3254
 |||||
 Db 98 TCTCAACTACTTCAGTATGTATACAAAGTTGGCTCCTCGTGGTATTTTATACAAAGTGGT 39
 |||||

the region [16,727].
 Plate: LSB0222 row: D column: 10.
 Location/Qualifiers
 1. .1156
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
 /note="vector: pCMV.SPORT6; Crown and developmental stages of spike formation in wheat cultivar Norstar. 4 mRNA populations were combined before constructing the library. The first mRNA population is from 1cm crown sections after 30 days of cold acclimation. The second is from 1cm crown sections after 11 days of deacclimation (before deacclimation plants were fully vernalized for 49 days). The third is from different developmental stages of spike formation (5 to 50mm) that still have not emerged from the leaf (dissection required). The last is from different developmental stages of spike and seed formation after having emerged from the leaf (visible). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 10.5%; Score 456.4; DB 7; Length 1156;
 Best Local Similarity 68.5%; Pred. No. 7.2e-129;
 Matches 659; Conservative 0; Mismatches 301; Indels 2; Gaps 2;

QY 3160 ATGTTTACTTGTGGGACCTGTACCAGTAATCTCAGTCTTCAGTATGTGCACA 3219
 DB 36 ATATTTTGTCTTGTGTATCTGGAAAGGAGGAGTTCAGAGTGGCTTACTG 95

QY 3220 AGATAGCTCCTCGTGAATCTACACTAGTGGCGGAGAAAGTTCCGGGTGGGCTGACAG 3279
 DB 96 AACTGTCTCTGTGTTTACAAAGTGGAAAGGAGGAGTTCAGAGTGGCTTACTG 155

QY 3280 CTTATGTAACGAAGATCAGAACTCGAGAGACGGTATGAGAGCGGAGCTTGGTTC 3339
 DB 156 CTTATGTTCTAAGGACCTGAACTGGTGAACCTGTTCTTGAAGTGGAGCACTGTGTT 215

QY 3340 TTAGTATCGTGGATATCTGTATGATGAGTTGACAAATGTCGATAATGCGCGAA 3399
 DB 216 TGAGTGACAAAGTGTGCTGTATGATGAGTTGATAAGATGCTGATAATGCGCGAA 275

QY 3400 GCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGTATCAATG 3459
 DB 276 GCATGCTGCATGAGGTGATGGAGCAGCACTGTATCCATTCGAAAGGCTGGAATTTATG 335

QY 3460 CTTGCTGAAACGCTCGGAGCTGTCTTGCATGTGCAATCTAGTGGTCCCGATACA 3519
 DB 336 CATCTTTGAATGCTAGGACATCTGTACTAGCATGTGCCAATCCATCTGAATCAGCTTACA 395

QY 3520 ATCGGCGCTTCTGTGATGATTAACATCAGCTTCTCCAACTCTACTTCTPAGATTTG 3579
 DB 396 ATCCAAAGGCTTCTGTGATGCAATATCCACTTCTCCAACTGCTGTCAAGGTTG 455

QY 3580 ATTTAATTTACTTAATGCTCGACAAACAGAGCAAGAAACGATCGTCTCGCCAGGC 3639
 DB 456 ACCTGATTTATCTGATCTTGGACAGGACAGCAACAACTGATAGACGCTGGCTAAGC 515

QY 3640 ATCTCGGCTTTACATCATATGAAACATGAGTTTCAAGAGGAGCGCTTATGATCTAC 3699
 DB 516 ATATTGTTTCATTTGCAATTCGAGAATCCAGAAGTAGTTGAGCACCGAGTCTTGATTTGC 575

QY 3700 AAACACTTACCGGTATATCACTATCTCGTCAGCATGTACATCTCTACATTTAAGTGATG 3759
 DB 576 CCACGTTAGTGTGGGTATCATGCTATGCAAGGAAGTACATTCAGCCAAAGTTATCTGATG 635

QY 3760 AAGCTGCTGAAGATTGATTAATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTG 3819
 DB 636 AAGCTGCAGAGAATTGACCGTGGCTACGTTGCAATGAGGCAAAAGGGGAAACAATCTG 695

QY 3820 GAAGCAGTAAAGGTGATTAACAGCCACACCTCGCAACTCGAAAGTATGATTCGTATCA 3879
 DB 696 GTAGCAGAAAGAGGTTCATCACAGCAACAGCTAGGCCAATTTGAGAGCTTGTATTCCTTA 755

QY 3880 GTCAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCGACAGAAG 3939
 DB 756 GTGAAGCACTGGCGCAAAATCTTTTTCAGAAAGTTGGTGGAGTGGAGATGTAATATCAAG 815

QY 3940 CTGTGGCCCTTTTAGAGCTGCTGTTTGCAGCAATCTGCTACTGATCATGCAACAGGTAGC- 3998
 DB 816 CTTTAGGCTTCTTGAAGTGCCTATGCAAGTCTTGAACCGATCATGCAACAGGACGA 875

QY 3999 ATAGACATGATCTTATCAGACTGAGTTCGGCCAGCGAGCGTATTTCGTCGGGCCAAC 4058
 DB 876 ATTGATATGATCTTATCATGACTGGGGTATC-CCAAGGTAAGGCAACGTAATATC 934

QY 4059 TTGCTAGCTCTCTGCGAGAGCTTATAGCAGATAAAATTTTCCCTGGCAGCTCTCTGGC 4118
 DB 935 TTGTTGGGGGCTATCCGAAACTTGTATGGAAGAAAGCCAGTTTGGGGGCCCTGTATGC 994

QY 4119 TT 4120
 DB 995 GT 996

RESULT 6
 DR503163
 LOCUS WS0298.B21_E02_SS-IB-A-FL-15 Picea sitchensis cDNA clone WS0298_E02
 DEFINITION 5, mRNA sequence.

ACCESSION DR503163
 VERSION DR503163.1 GI:70241689
 KEYWORDS EST.

SOURCE Picea sitchensis (Sitka spruce)
 ORGANISM Picea sitchensis
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE 1 (bases 1 to 907)
 AUTHORS Ralph,S., Kolosova,N., Oddy,C., Cooper,D., Butterfield,Y.,
 Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
 Babakoff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
 Mayo,M., Moran,J., Olson,T., Wong,D., Friedmann,M.F., Ritland,C.E.,
 Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C.,
 Ritland,K. and Bohlmann,J.

TITLE The spruce transcriptome: Analysis of expressed sequence tags from
 multiple cDNA libraries

JOURNAL Unpublished (2005)
 COMMENT Contact: Joerg Bohlmann
 Genome BC forest genomics program
 University of British Columbia
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
 Vancouver, British Columbia, Canada, V6T 1Z3

Fax: 1-604-822-2114
 Email: bohlmann@msl.ubc.ca
 Plate: WS0298 row: E column: 02
 High quality sequence stop: 907.

FEATURES
 Location/Qualifiers
 1. .907
 /organism="Picea sitchensis"
 /mol_type="mRNA"
 /cultivar="PB3-425"
 /db_xref="taxon:3332"
 /clone="WS0298_E02"
 /sex="Hermaphrodite"
 /lab_host="E. coli DH10B cells"
 /clone_lib="SS-IB-A-FL-15"
 /notes="Organ: Bark (with phloem and cambium attached) from
 two year old clonal trees grown under greenhouse
 conditions in standard potting soil mixture; Vector:
 pBluescript II SK (+) XR; Site 1: XhoI (5' end of cDNA);
 Site 2: BamHI (3' end of cDNA); Adult white pine weevils
 (Pissodes strobi) were caged with mesh bags on sapling

trees. Bark tissue with phloem and cambium attached was harvested 2 hours, 6 hours and 48 hours after continuous feeding by the insects. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 20 micrograms of mRNA according to the full-length cDNA library construction method described by Carninci P. et al. (2000), Genome Research 10(10):1617-1630 and directionally ligated into the pBluescript II SK (+) vector digested with XhoI (5' end) and BamHI (3' end). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation."

ORIGIN

Query Match	10.1%	Score 439.4;	DB 8;	Length 907;
Best Local Similarity	67.8%	Pred. No. 1.2e-123;		
Matches 614; Conservative	0;	Mismatches 291;	Indels	0

2178	QY	AAGTACATCAGATCATAGAGGAGACGTGTGGAGCGTGGAGGATATCTTAAATATCGAC	2237
3	Db		
3	Db	AAATACATCAGGATTTTATAGAGCAGACAATGGAAGTAGAAGCGGATCTTTTGAACGTTGAT	62
2238	QY	ATGTTCAGACATTTATGACCACATGATCCTGATCTATACGCCAAAAATGTTTCATATCCCACTC	2297
63	Db		
63	Db	GCAATGAAATCTTGGATTTATGATGAGGACCTTTATGGGAAAATGGTTAGATACCCCTC	122
2298	QY	GACATCATCCCCCTGTTTGGACACTGAGTGTTCAGGAAGTTCTACTCTTTTACTTACCAACG	2357
123	Db	GAAGTTCTTGCCATTTTGTGATATGTTTGTGATGATGTGCTGTGCGCTTTAAACCTTTG	182
2358	QY	TTTGTAGAGACATATTGAGGCGCAGACCTTTCAATCTCAAGCATCGGTGACATCGGTGAA	2417
183	Db	TGGGAGAAGCACATACAGGCAAGAATATTCAATCTCAAGATTCAGTTTAACTATCGCAGAA	242
2418	QY	CTCAACCCCTTCAGATATAGACAAATTTGGTTTCTGTTTAAAGGAATGGTTATCCCGTGCAGT	2477
243	Db	CTGATCCATCAGATTTTGAAGAAATGGTTTTAGTGAAGGATGATCATACATGACGC	302
2478	QY	TCATCATACTGAAATTAAGGGGGCTCTTCAAAATGTTTAGTGTGTGTCACTCGCCT	2537
303	Db	TCAAATAACTGAAATCAAAGAGGCATTTTTCAGATGCCCTGTTGTTGGCTATGTCCA	362
2538	QY	CCGCTAGTTACAGTTGTTAAAGGGCGGTGAGGAGCCACAGGTTGTGAAGAGCCAGAA	2597
363	Db	CCACCAGTGCCTGTTGATAGAGTGCAAATTTGATGAACCTTACTCGATGTGGAAGGCCAGAA	422
2598	QY	TGTGCACGACGGHAATGCTATGTCCTTATTCACAATCGATGCATCTTTTGCAAAATAAGCAG	2657
423	Db	TGTGCTGCTATGAATCCATGACTTTTAGTTTCAATATAGATGCAGGTTTGTGTGCAACACAG	482
2658	QY	ATAGTGGGTCTTCAAGAACTCAGATGCCATTTCTTGAAGGAGAGACTCCACACAGTC	2717
483	Db	ATCGTAAGGCGTCAAGAAACCCCTGATGCTATTTCAGAGGGGAGAAACCCACATACAGTC	542
2718	QY	AGCATGTGTTTATACAACTATGTTTGTATGCTGTGAAGCTCGAGATCGTATTGAGGTA	2777
543	Db	AGCTTGTTAATGATGACAAACTAGTTGATGCTGCCAAGCCTGGTGACAGTTGAGGTA	602
2778	QY	ACAGGAGTTTTCAGGCCATGGCAGTTTCGAGTTTGGTCCGAATCAACGAAATTTACGAGCA	2837
603	Db	ACAGGAATCTTCAGGGCCATGAGTGTTTAGAGTAGGACCAAAACCAGAGGACAGTCAAAATCC	662
2838	QY	TTGTATAAGACCTTACATCGATTCGGTGCAGCTCAAGAAGTCTGACAGAGGGGTTCAGTCGAA	2897
663	Db	CTCTTTAAGACCTTCAATTGATCGCTCTTCACTTGAAGAAGGCAGACAAGACACAGCTGCAA	722
2898	QY	ACTGAAGATCCTATGGAGATGGAATAAGGAGAATGATATGTATGCTGGGTATCATGAGAAAT	2957
723	Db	GCTGAAGATCCAAATGAGCGGTGATGCCATAGTGAACAGCTTTGCTTCTTTTCAAGAGGGT	782
2958	QY	GATACTTCGAAGACTGCTAATGAAGCAAGAATTCAAAAAATTTAAGAGCTGTGCCAGCTC	3017
783	Db	GATATTTTCACCCGTTCAGTATGAAGCCAGATTTGAGAAGTTTGAAGAAGCTGTCAAGGAG	842

Qy	3018	CCGGGCATTTATGATAGACTTTCAAGTCTGCTCGCTCCCAAGCATTTTGGGAGCTTGAAGAT	3077
Db	843	CCTGATATATGAGAAGTTAAACCAAGTCATTGGCTCCAAGTATTTTGGGAGCTGGAAGAT	902
Qy	3078	ATTAA	3082
Db	903	GTGAA	907

RESULT 7

BF051102	BF051102	770 bp	linear	EST 18-MAY-2001
LOCUS	EST436277	tomato developing/immature green fruit Lycopersicon		
DEFINITION	esculentum cDNA clone CLEM21H9 5' sequence, mRNA sequence.			
ACCESSION	BF051102			
VERSION	BF051102.1			
KEYWORDS	GI:10804998			
	EST.			

REFERENCE
SOURCE

ORGANISM
Lycopersicon esculentum
Lycopersicon esculentum (Solanum lycopersicum)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE

REFERENCES
1. (cases 1 to 70)
AICALA, J., VEBALOV, J., WHITE, R., van der Hoeven, R. S., Holt, I. E.,
LIANG, F., HANSEN, T. S., Craven, M. B., Bowman, C. L., Ronning, C. M.,
NIERMAN, W., FRASER, C. M., Martin, G. B., GIOVANNONI, J. J. and
TANKSLEY, S. D.

TITLE

JOURNAL COMMENT

Generation of ESis from tomato fruit tissue, immature green
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.

304102

```

/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEW21H9"
/tissue_type="fruit"
/dev_stage="immature green (5-35 days post-anthesis)"
/lab_host="SOLr"
/clone_lib="tomato developing/immature green fruit"
/note="Vector: pBluescriptSKmClnadap; Site 1: EcoR1; Site 2: Xho1; Fruit were tagged at 5 dpa (0.5 cm) and harvested at 7 day intervals through 35 dpa. Equal masses of tissue from each stage were combined (including seeds and locules) prior to mRNA isolation."

```

ORIGIN

Query Match	10.0%	Score 434.2;	DB 2;	Length 770;
Best Local Similarity	73.5%;	Pred. No. 4.6e-122;		
Matches 567;	Conservative	0;	Mismatches 203;	Indels 1;
			Gaps	1;

Qy	2996	ACTTAAAGAGCTGTCCAAAGCTCCCGGCATTATGATAGACTTTCAAGGTCGTGGCTCC	3055
Db	1	ATTGATAGAACTATCTTAAACAGCCCTGATATCTATGAAAGGCTGACCAGGTCCTTGGCACCC	60

Qy	3056	AAGCATTTGGAGCTTGAAGTATTTAAAAAGGGTCTTTTTCAGCTCTTTGGTGGAA	3115
Db	61	AAACATATGGGAGTTGGATGATGTTAAAGAAAGGCCCTTTTTCAGCTCTTTGGTGGAA	120

3116 GGCTAAGAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTGTTGG 3175
121 TGCAATTGACCTTGCCTCTGGAGCAAGCTTCCGTGGTGAATCAACATTCCTCTTGTGG 180

Qy	3176	GGACCTCGTGACCAAGTAAATCTCAGCTGCTTCAGTATGTGCACAGATAGCTCCTCGTGG	3235
Db	181	TGATCGGGGACCGCAATACAACCTGCTCAGTATATTCACAACTGTCTCCTCGTGG	240

Thu Dec 8 10:36:56 2005

```

/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/note="Vector: pBluescriptSMCUadapt; Site 1: EcoRI;
Site 2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."

```

ORIGIN

Query Match	10.0%;	Score 433;	DB 3;	Length 788;
Best Local Similarity	72.0%;	Pred. No. 1.1e-121;	Indels 0;	Gaps 0;
Matches 565;	Conservative	0;	Mismatches 220;	
QY	3321	GAGAGCGGAGCTTGGTTCCTTAGTGATCGTGGATATGCTGTATCGATGAGTTCGACAA	3380	
Db	2	GAGAGTGGCGCTCTGGTCTTTGAGTGACAGAGGATCTGCTGTATTGACGAGTTTGACAA	61	
QY	3381	ATGCTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTA	3440	
Db	62	ATGCTCTGACAGTCCAGAGGATGTTTACATGAGGTGATGAGGCAACAAACTGTTTCAAT	121	
QY	3441	GCCAAAGGGGGTATCATTTGCTCGTGAAGCTCGGACGCTCTCTCTTCCTTGCATGTGCAAT	3500	
Db	122	GCAAGGCTGGAATATTCGCTTCGCTTAATGCTAGGACTTCAGTATTGGCATGTGCAAT	181	
QY	3501	CTTAGTGGGTCCGATACATGCGCGCTTTCTGTGATTGATTAACATCCAGCTTCCTCCA	3560	
Db	182	CCAATTGGCTCCGCTTACAAATCCCGGATTAATCTGCTGATTAATATACACCTTCCACCT	241	
QY	3561	ACTCTACTTTCTAGATTGATTTAAATTTACTTAATGCTCGACAAACAGAGCAACAAAC	3620	
Db	242	ACCTGCTGCTAGATTGATTTGATATATTAATTTCTAGACAAAGCAGATGAGCAGACC	301	
QY	3621	GATCGTCTGCTCGCCAGGCAATCTCGTGGCTTTACACTATGAAACTATGAAGTTTCAAAG	3680	
Db	302	GACAGGCGCTTGCACAAAGCACATAGTTGCTTTACACTCTGAGAAATCTTGAGACTCTGAG	361	
QY	3681	CAGAGCGCTTAGATCTCAACAACTTACCCGCTATATACCTATGCTTCGTCGATGTA	3740	
Db	362	CAAGAAGTGATTTGACCTTCCAACTTACCTTCACTATTGATGCTATGCTCGAAACATATA	421	
QY	3741	CATCTCATTTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGGGC	3800	
Db	422	CATCCACAATTTATCTGATGAAGCAGCTGAAGATTTGACTAGAGGTTATGTTGAGATGAGA	481	
QY	3801	CAAAAGGGCAACTTTCTCGGAGCAAGTAAAGGTTAAACAGCCACACCTCGGCAATC	3860	
Db	482	AGAAAAGGGAAATTTCCCGGTAGCAGTAAAGGTTGATTTACAGCTACACCAAGGCACTA	541	
QY	3861	GAAGATGATTTGCTGATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAG	3920	
Db	542	GAGAGTTTGATACGCTTAGAGTGAAGGCTTTGCTCGATGCGTTTTCAGAAAAGGTGGAG	601	
QY	3921	AAAGTTGATGACAGCAAGAGCTGTCGCGCTTTTAGAGCTGCTTTTGCAGCAATCTGCTACT	3980	
Db	602	AAAAGAGATGTTAGTGAGGCTTTTCGACTTCTAGAAGTTGCTATGTCAGCAGCTGCAACT	661	
QY	3981	GATCATGCAACAGGTACGATAGCATGGAATCTTATCAAGCTGAGTGTTCGGCCAGCGAG	4040	
Db	662	GACCATGCTACAGGAACCATGACATGGATCTCATCAACCTGAGTATCTGCAAGTGAA	721	
QY	4041	CGTATTCGTCGGGCCAACTTGTGCTGCTCGGAGAGCTTATAGCAGATAAAATTTCA	4100	
Db	722	AGGATGAGAGGGAGAAATTTGGTGTCAAGCACCCGCAACATAATTTATGGAGAGATCGAG	781	
QY	4101	CCTGG 4105		
Db	782	CTTGG 786		

QY	3236	AATCTACACTAGTGGCGAGGAAGTTTGGCGGTGGCTGACAGCGTATGTAACGAAGA	3295	
Db	241	TATATACACAAGTGACAGGAAGTTTCACTGTGGGTTGACTGTATGTAGCCAAAGA	300	
QY	3296	TCAGAACTCGAGACACGATTTGAGAGCGGAGCTTTGGTCTTCTAGTGATCGTGGGAT	3355	
Db	301	TCCGAGACTGGTGAACCTGCTTCTGAGAGTGCGCTCTGGTCTTCTGAGTGACAGAGGAT	360	
QY	3356	ATGCTCTGATCGATGCTGACAAAATGCTGATAATGCCCGAAGCATGCTTCATGAGGT	3415	
Db	361	CTGCTGTATGACAGTTTGACAAATGCTGACAGTGCAGGAGCATGTTACATGAGT	420	
QY	3416	AATGGAGCAACAAACGGTATCTTGTGCAAGGGGGTATCATGCTCTGCTGACACCTCG	3475	
Db	421	GATGGAGCAACAACTGTTTCAATTTGCAAGGCTGGAATTTATCGTTCGCTTAATGCTAG	480	
QY	3476	GAGCTCTGCTTGCATGTGCAAACTCTAGTGGGTCCCGATACAATGCGCGCTTTCTGT	3535	
Db	481	GACTTCAGTATGCGCATGTGCAAACTCAATTTGGCTCCCGTTACATCCCGCATATCTGT	540	
QY	3536	GATTGATAACATCCAGCTTCTTCCAACTCTACTTCTAGATTTGATTTAATTTACTTAAT	3595	
Db	541	CATTGATAATATACACTTCCACCTACCTGCTCTGATGTTGATTTGATATATTAAT	600	
QY	3596	GCTCGACAAACAGCAGCAACATCGTCTGCGCAGGAGATCTCGTGGCTTTTACA	3655	
Db	601	TCATGACAAAGCAGTACGAGCAGCGGCTTGCAGGAGCAGATGTTGCTTTTACA	660	
QY	3656	CTATGAAACTATGAAGTTTCAAGAGCAGGAGCTTATGATCTACAAACACTTACCGGTA	3715	
Db	661	CTCTGAGAATCCTGAGAACTCTGAGCAAGAGTGAATGACCTTCCAACTTAGCTTATTA	720	
QY	3716	TATCACTATGCTGTCAGCATGATCATCTTACATTAAGTATGATGAAGCTGC	3766	
Db	721	CTTGAGCTATGCTCG-AAACATATATACATCCAAATTTATCTGATGAGCAGC	770	
RESULT 8				
LOCUS				
DEFINITION				
EST586433 tomato breaker fruit Lycopersicon esculentum cDNA clone				
CLEG59A13 5' end, mRNA sequence.				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Lycopersicon esculentum (Solanum lycopersicum)				
Lycopersicon esculentum				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;				
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.				
1 (bases 1 to 788)				
Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A.,				
Tsai, J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S.,				
Romling, C.M., Fraser, C.M., Martin, G.B., Fankley, S.D. and				
Giovannoni, J.				
Generation of ESTs from tomato fruit tissue, breaker stage (2002)				
Unpublished (2002)				
Contact: CUGI				
Clemson University Genomics Institute				
Clemson University				
100 Jordan Hall, Clemson, SC 29634, USA				
Email: http://www.genome.clemson.edu/orders/index.html				
This clone is available through the Clemson University Genomics				
Institute				
Seq primer: T3.				
Location/Qualifiers				
1..788				
/organism="Lycopersicon esculentum"				
/mol_type="mRNA"				
/cultivar="TA496"				
/db_xref="taxon:4081"				
/clone="CLEG59A13"				
/tissue_type="pericarp"				
FEATURES				
source				

```
RESULT 9
AY408033      2522 bp   DNA      linear      GSS 12-DEC-2003
LOCUS      Homo sapiens HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION      genomic survey sequence.
ACCESSION      AY408033
VERSION      AY408033.1   GI:39764004
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 2522)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL      Science 302 (5652), 1960-1963 (2003)
PUBMED      14671302
REFERENCE      2 (bases 1 to 2522)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
TITLE      Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
JOURNAL      Rockville, MD 20850, USA
COMMENT      This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source      Location/Qualifiers
1..2522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2522
/locus_tag="HCM3089"
ORIGIN
Query Match      9.9%; Score 429.2; DB 10; Length 2522;
Best Local Similarity 53.6%; Pred. No. 2.6e-120;
Matches 1113; Conservative 0; Mismatches 878; Indels 87; Gaps 7;
QY      2181 TACATCCAGATCATAGAGGAGCTGGAGCGTGGAGGATCTCTTAATTCGACATG 2240
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
507 TACATGCAACGACTTGGGAGATTAATGTTATGGTGAGCCATTTTAAATGGAACGT 566
QY      2241 TCAGACATTTATGACCATGCTCTGATCTATACGCAAAATTTGTCGATACCCACTCGAC 2300
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
567 GAACACATCAATCATTTGACAAAATTTGTACAGACAACTCATCTTTACCCACAGGAA 626
QY      2301 ATCATCCCCCTGTGTGACACTGAGTCTCAGGAA-----GTTGCTACCTTTTACTACCA 2354
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 GTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTCTTTGACCGTTTACCCTGACTCA 686
QY      2355 ACGTTTGAGAGCATATTGAGGCCACACCTTTCAATCTCAAGCATCGGTGACATCGGT 2414
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
687 ATCTTAGAACATCAGATTCAAGTAAGAACCAATTCACGCCATTGAAGCAATGAATATGGA 746
QY      2415 GAACTCAACCTTCAGATATAGCAAAATGTTCTGTTTAAAGGAATGGTTATCCGGTGC 2474
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
747 AACCTGAATCCAGAACATTTGACACAGCTCATCACCATCAGCGCATGTTGATCAGACA 806
QY      2475 AGTTCTPATCATACCTGAAATTAAGGGGGCTTCTTCAAAATGTTTAGTGTGTGCTACTCG 2534
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
807 TCCAGCTGATTTCCCGAGATGACAGGAGGCTTCTTCCAGTGCCCAAGTGTGTGCCACACG 866
QY      2535 CTTCCGCTAGTTACAGTTGTTAAAGGGCGGTGTGAGGAGCCCAAGAGGTGTGAAAGCCA 2594
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
867 ACCCGGGTGAGA---TGGACCGCGCGCATTTGCAGAGGCCACAGTGTGTGCGGGCG- 919
12595 GAATGTCGAGCAGCGGAATGCTATGCTCTTATTTCAATTCGATCGACATTTTCAATAAAG 2654
12600 --CTGCCACACCCACACAGCATGCACTCATCAACACCGCTCCCTCTTCTCTGACAAAG 2655
2655 CAGATAGTGCCTCTTCAAGAACTCCAGATGCCATTTCTTGAAGGAGAGACTCCACACACA 2714
2660 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
978 CAGATGATCAGCTTCAGAGTCTCCGGAAGACATGCCCTGACAGGAGACACACACACA 1037
2715 GTTCAGCATGTGTTTATACAACACTATGCTGTGTAAGCCCTGGAGATCGTATTTGAG 2774
1038 GTTATCTCTTGTCTCACAATGATCTCGTTGACAAGGTCACGCTGGGGACAGAGTGAAT 1097
2775 GTTACAGAGATTTTCAAGCCCATGCGAGTTCGAGTTGGTCCGAATCAACAGCAATTACCA 2834
1098 GTTACAGGATCTATCGAGCTGTGCTTATTCGAGTCAATCCAAAGAGTGTGATTAATGTGA 1157
2835 GCATTTGATAGACCTTACATCGATTCGCTGACAGTCAAGAGTCTTGACAGGGGTGCAGCTG 2894
1158 TCTGTCTACAAACCCACATGATGTCATTTATCGGAAACCGATCGAAACATCTCTG 1217
2895 CAAACTGGAAGATCTTATGAGATGGAATGAAGAGAAATGATATGCTGCTGGTATCATGAA 2954
1218 CA-----TGSCCTTGTATGAA 1232
2955 AGTGATATCTTCAAGAGCTGCTTAATGAAGCAAGATTCAAAACTTAAAGAGCTGTCCAAG 3014
1233 GAAGCAGAAACAGAAACTTTTTTCAGAGAAACGCTGTGGAATTTCTTAAAGGAATCTTCCAG 1292
3015 CTCCTGGGATTTATGATAGACTTTTCAAGTCTGCTGCTCCAAAGCATTTGGGAGCTTGAA 3074
1293 AAACAGACATTTATGAGAGGCTTGTTCAGCTTGTGCTTCCAGGATTTATGAACATGAA 1352
3075 GATATTTAAAAAGGCTCTTCTTGGCAGCTCTTTGGTGGGAAGGCTTAAGAAAAATTT----- 3128
1353 GATATAAGAGGGAATTTTGTCTTGTGCTTGTGCGGAGCAAGGAAGATTTTGTAGTCAAC 1412
3129 CATCTGGAGCATCTTTTCGAGGTGACATCAATGTTTACTTGTGTTGGGAGCCCTGTGTACC 3188
1413 ACTGGAAGGGGCAAAATTTTGGGCTGAGATCAACATCTTGTGTGTGCGGACCTGTGTACC 1472
3189 AGTAAATCTCAGCTGCTTCAAGTATGTCACAGATAGCTCTCTCGTGAATCTACACTAGT 3248
1473 AGCAAGTCCAGCTGCTGAGTACGTTGTAACAACCTGCTCCCGAGGGGCCAGTACAGTCT 1532
3249 GGGCAGGAGGAATTCGGCGGTTCGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
1533 GGGAGGGCTCCAGTGCAGTGGCTCTCTGCTGAGTGACAAAGGATCTGCTGATTCGAT 1592
3309 GAGACGGTATTCGAGAGCGGAGCTTTGGTCTTGTAGTATCGTGGGATATGCTGATTCGAT 3368
1593 CAGCTGGTCTCGACAGAGGTGCTCTTGTCTGAGTGACAAAGGATCTGCTGATTCGAT 1652
3369 GAGTTGCAAAAATGCTGATTAATGCGGAGCATGCTTTCATGAGTATGAGGACAA 3428
1653 GAGTTGCAAGATGAATGAAGATGTAAGATCGGATTCATGATGAAGTCAATGGAACAGCAG 1712
3429 ACGGTATCTGTAGCAAAAGGGGATCATTTGCTGCTGCAACGCTCGGAGCTGTGCTCTT 3488
1713 ACTGTCTCATTTGCAAGGCTGGGATCATCTGTGAGTCAATGCGGCGACCTCTGTCTCTG 1772
3489 GCATGTGCAAACTCTAGTGGTCCCGATCAATGCGCGGCTTTTGTGTGATGTAATACATC 3548
1773 GCAGCAGCAAAATCCCATTTGAGTCTCAGTGAATCTTAAAAAACAACCACTTGAACATC 1832
3549 CAGCTTCTCCAACTCTACTTCTAGATTTGATTTAAATTTACTTAAATGCTGCAACAA 3608
1833 CAGCTGCTCTCATCTTATTTATCAAGGTTTGAATTTCTCTCTCATGCTGGAACCTCAG 1892
3609 GACGAGCAAAAGCATGCTGCTCGCCAGGATCTCTGCTGGCTTTTACACTATGAAAACTAT 3668
1893 GACGAGCCATGACAGGCGCTCTGGCTCACCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1952
3669 GAAATTTTCAAGCAGGAGCGCTTATGATCTTACAAACACTTACCGCGTATATCACTATGCT 3728
```

1953	Db	GAGCAGCAGGAGGAGCTCTGGACATGGCGTGTAAAGGACTACATTGCCTACGGG	2012
3729	Qy	CGTCAGCATGTACATCCTACATTAATGATGTAAGCTGCTCAAGATTGTATTATGCTAT	3788
2013	Db	CACAGCACCATCATGCCGCGCTAAGTGTAGGAAGCCAGCAGGCTCTCATCGAGGCTTAT	2072
3789	Qy	GTTGAGATGCGCCAAAGGGCACTTTCCTGGAGCAGTAAAAAGGTGATAACAGCCACA	3848
2073	Db	GTAGACATGGAAGATGTGGCAGTAGCCGGGAA-----TGGTTTCTGCATAC	2120
3849	Qy	CCTCGGCAACTCGAAAGATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTCT	3908
2121	Db	CCTCGACAGCTAGAGTCAATTAATCCGTTAGCAGAGCCCATGCTAAAGTAAGATTGCT	2180
3909	Qy	GAAGTGGTAGAGAAAGTTGATGTCAGCAGAAAGCTGTGCGCTTTTAGACGTGCTTTGCAG	3968
2181	Db	AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCCTCCATCGGAAGCTCTGAAG	2240
3969	Qy	CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCAGCACTGGAGTG	4028
2241	Db	CAGTCTGCAACTGATCCCGGACTGGCATCGTGGACATATCTATCTTCTACTACGGGGATG	2300
4029	Qy	TCGGCCACGAGCGTATTCGTCGGGCCCAACTTGTAGTGTCTCTGCAGAGCTTATAGCA	4088
2301	Db	AGTGCCACCTCTCGTAAACGGAAAGAGAATTAGCTGAAGCATTTGAAAAGCTTATTTTA	2360
4089	Qy	GATAAATTTACCTGGCAGCTCTCTGCGTTGAAGACCAGTCAGCTTCTTGAGGATATC	4148
2361	Db	TCCTAA-----GGGCAAAACACCAAGCTCTAAATACCAAGCACTTTTGAAGATATT	2411
4149	Qy	CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTC	4208
2412	Db	CGGGACAACTCTGACATAGCAATTACTAAAGATAGTTTGAAGAGCACTCGTGCCCTG	2471
4209	Qy	CAAGGAGAGGCTTTTCTACTGTCTCATGGTGACATAGT	4246
2472	Db	GCAGATGATGATTTCTGACAGTGTACTTCTGACAGTGTCTGGGAAGACCGT	2509

RESULT 10	BM410852	775 bp	mRNA	linear	EST 22-JAN-2002
LOCUS	EST585179	tomato breaker fruit Lycopersicon esculentum	cDNA clone		
DEFINITION	CLEGS4H12	5' end, mRNA sequence.			
ACCESSION	BM410852				
VERSION	BM410852.1	GI:18262482			
KEYWORDS	EST.				
SOURCE	Lycopersicon esculentum (Solanum lycopersicum)				
ORGANISM	Lycopersicon esculentum				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanales; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 775)				
AUTHORS	Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai, J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S., Ronning, C.M., Frazer, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.				
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage (2002)				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: CUGI				
	Clemson University Genomics Institute				
	Clemson University				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Email: http://www.genome.clemson.edu/orders/index.html				
	This clone is available through the Clemson University Genomics Institute				
	Seq primer: T3.				

```

FEATURES
  source
    seq primer: 33
    Location/Qualifiers
      1..775
        /organism="Lycopersicon esculentum"
        /mol_type="mRNA"
        /cultivar="TA496"

```

```

/db_xref="taxon:4081"
/clone="CLEG54H12"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/note="Vector: pBluescriptSMCuaadapt; Site_1: EcoRI;
Site 2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."

```

ORIGIN	Query Match	9.8%	Score 428	DB 3	Length 775
	Best Local Similarity	73.6%	Pred. No. 3.9e-120		
	Matches 545	Conservative	0	Mismatches 195	Indels 0
				Gaps 0	
QY	3321	GAGAGCGGAGCTTTGGTCTTTAGTGATCGTGGGATATGCTGTATCGATGAGTTTCGACAA	3380		
DB	2	GAGAGTGGCGCTCTGGTCTTGGTGACAGAGGGATCTGCTGTATTCGAGTTTGACAA	61		
QY	3381	ATGTCTCATATATGCCGAAGCATGCTTTCATGAGTAATGGAGCAACAAACGGTATCTGTGA	3440		
DB	62	ATGTCTGACAGTGCCAGGAGCATGTTACATGAGTGTATGGAGCAACAAACGTGTTCAAT	121		
QY	3441	GCCAAAGGGGGTATCATTTGCTCTGCTGAAACGCTCGGACGTCTGCTTGGCATGTGCAAA	3500		
DB	122	GCAAAGGCTGGAATTAATCGCTTCGCTTAATGCTAGGACTTCAGTATTTGGCATGTGCAAA	181		
QY	3501	CCTAGTCGGTCCCGATACAATGCGCGCTTTCTGTGATTCATAACATCCAGCTTCCTCCA	3560		
DB	182	CCAATTGGCTCCCGTTACAATCCCGGATTAATCTGTCAATGATTAATACACCTTCCACCT	241		
QY	3561	ACTCTACTTTCTAGATTTGATTTTAATTTACTTAATGCTCGACAAACAGACGAGCAAAAC	3620		
DB	242	ACCTCTGCTGTAGATTTGATTTGATGATATATTTAATCTAGACAAAGCAGATGACAGACC	301		
QY	3621	GATCGTGTCTCGCAGGCACTCTCGTGGCTTTACACTATGAAACTATGAATTTCAAG	3680		
DB	302	GACAGGCGCCTTGCAAGCACATAGTTGCTTTACACTCTGAGAATCTCTGAGAACTCTGAG	361		
QY	3681	CAGAGCGCTTAGATCTATCAAACTTACCCGCTATATCACTATGCTCGTCAGCATGTA	3740		
DB	362	CAAGAGTGATTTGACCTTCCAACTTAGCTTCATCTTGAGCTATGCTCGAAAAACATATA	421		
QY	3741	CATCTCTAATTAAGTGATGAAGCTGCTGGAAGATTTGATTAATGGCTATGTTGAGATGGC	3800		
DB	422	CATCCACAATTAATCTGATGAAGCAGCTGAAGAAATTTGACTAGAGGTTATGTGGAGATGAGA	481		
QY	3801	CAAAAGGCAACTTCTCGGAAGCAGATGAAGGTGATAACAGCACACCTCGGCNACTC	3860		
DB	482	AGAAAGGGAAATTTCCCGGTAGCAGTAAAGGTGATTAACAGTACACCAAGGCGACTA	541		
QY	3861	GAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAG	3920		
DB	542	GAGAGTTTGATACGCTTAGGTGAAGGTCTTGCTCGATGCGTTTTTCAGAAAAGGTGGAG	601		
QY	3921	AAAGTTGATGACGAGAAGCTGTGGCGCTTTTAGACGTGCGTTTTCGACGAATCTGCTACT	3980		
DB	602	AAAAGAGATGTAGTGGAGGCTTTTTCGACTTCTAGAAAGTTGCATTCGAGGAGTCTCGAACT	661		
QY	3981	GATCATGCAACAGGTACCATAGACATGGATCTTATCAGCACTGGAGTGTCCGCCACGCGAG	4041		
DB	662	GACCATGCTACAGGAACCAATTGACATGGATCTCATCAACAATGGAGTATCTGCAAGTGA	721		
QY	4041	CGTATTTCTGTCGGGCCCAACTT	4060		
DB	722	AGGATGAGAAGCGGAGAATT	741		

RESULT 11

AY408034 2357 bp DNA linear GSS 12-DEC-2003
LOCUS Pan troglodytes HWM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY408034
VERSION
KEYWORDS AY408034.1 GI:39764005
GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan. (bases 1 to 2357)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferrier, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Infering nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 2357)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferrier, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1. .2357
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1. .>2357
/locus_tag="HWM3089"

Query Match 9.4%; Score 408.4; DB 10; Length 2357;
Best Local Similarity 50.4%; Pred. No. 7.1e-114;
Matches 1015; Conservative 0; Mismatches 922; Indels 78; Gaps 6;

QY 2181 TACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATCTCTAAATATCGACATG 2240
DB 342 TACATGCAAGACTTGGGGAGATTAATGTTATTTGGTGAGCCATTTTAAATGTGAATGT 401
QY 2241 TCAGACATTTATGACCATGATCTCTATACGCAAAAATTTGTCGATACCCACTCGAC 2300
DB 402 GAACACATCAATCATTTGACAAAATTTGTACAGACAACTCATCTCTTACCACAGGAA 461
QY 2301 ATCATCCCTCTGTGACACTGAGTGTGAGAA-----GTTGCTACTCTTTACTACCA 2354
DB 462 GTTATTTCCAACTTTTGACATGGCTGTCAATGAAATCTTCTTACCGCTTACCCTGACTCA 521
QY 2355 ACCTTTGAGAGCATATTGAGCGACACCTTTCAATCTCAAGCATCGGTGCGACATCGT 2414
DB 522 ATCTTAGAATCATGATTTCAAGTAAGACCAATTCACGCAATGAAGCAATAAGATATGAGA 581
QY 2415 GAACTCAACCTTTGAGATATAGCAAAATTTGTTCTGTAAAGGAATGTTATCCGGTGC 2474
DB 582 AACCTGATCCAGAGCATTTGACCACTCATCACCATCATGTCAGTGGCATGGTATCAGGACA 641
QY 2475 AGTTCTATCATCTGAAATTAAGGGGCGCTTCTTCAAAATGTTTGTGTTGTGCTACTCG 2534
DB 642 TCCAGCTGATTTCCGAGATGCGAGGCGCTTCTTCCAGTGCCAAAGTGTGTGCCACACG 701
QY 2535 CTCCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCAACAAGGTGTGAAAGCCA 2594
DB 702 ACCCGGTGAGGA---TGGACCGCGCGCAATTTGACAGAGCCCAATGTGTGCG-----GG 752
QY 2595 GAATGTGCAGCAGGAATGCTATGCTCTCTATTTCACAATCGATCGACTTTTGGCAATPAG 2654

753 CGTGCACACCAACCACAGCATGGCACTCATCCACAACCGTCCCTCTCTCTGCAAG 812
2655 CAGATAGTCGGTCTTCAAGAAAATCCAGATGCAATTCCTGAAGGAGAGACTCCACACACA 2714
813 CAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCTGCAGGCGAGACACACACA 872
2715 GTCAGCATGTGTTTATACAACTATGTTGTGATGCTGTGAAGCTCGAGATCTGATATGAG 2774
873 GTTATCTCTGCTCACAATGATCTCTGTGCAAGGTCGAGCTCCAGCTGGGACAGAGTGA 932
2775 GTAAACAGGAGTCTTCAAGGCCATGGCAGTTCCAGTTGTCGGAATCAACGAACATACGA 2834
933 GTTACAGGCAATCTATCGAGCTGTGCTTATTCGAGTCAATCCAAAGATGAGTATGGAAG 992
2835 GCATTTATAGAACCCTACATCGATTCGTCGTCAGCTCAAGAAGTCTGACAGGGGTGCACTG 2894
993 TCTGTCTACAAAACCCACATTTGATGTCATTTATTCGGAARACGATGCAAAACGTCG 1052
2895 CAAACTGAAAGATCCTATGAGATGGATAAGGAGAATGATATGATGCTGGGTATCATGAA 2954
1053 CA-----TGGCCTTGATGA 1067
2955 AGTGATCTTCAGAAGCTGCTATTAAGCAAGAAATTCAAAACCTTAAGAGCTGTCCAAG 3014
1068 GAAGCAGAACAGAAACTTTTTTTCAGAGAAACGCTGTGGAATTTGCTTAAGGAACCTTCCAGG 1127
3015 CTCCCGGCGATTTATGATAGACTTTCAAGGTCTGCTGCTCCAAAGCATTTTGGGAGCTTGAA 3074
1128 AAACAGACATTTATGAGAGGCTTGTCTCAGCCTTGGCTCCAGCATTTATGAACATGAA 1187
3075 GATATTAAGAGGCTCTTTTTCAGCTCTTTGGTGGGAAGCTTAAGAAATTT----- 3128
1188 GATATAAGAGGGAATTTTGTCTTTCAGCTCTTTGGCGGGAAGAGGATTTTAGTCAC 1247
3129 CCATCTGGAGCATCTTTCCAGGTGACATCAATGTTTACTTCTTGGGAGCCCTGGTACC 3188
1248 ACTGGAAGGGGCAAAATTTCCGGCTGAGATCAACATCTTGTGTGCGGACCTTGGTACC 1307
3189 AGTAAATCTCAGCTGCTTCAGTATGTGCAAGATAGCTCTCTCGTGAATCTACACTAGT 3248
1308 AGCACTCCAGCTGCTGAGTATGTACAACCTGTCGCCAGGGGCCAGTACAGTCT 1367
3249 GGGCAGGAGGAGTTCGGCGGTGGGCTGACAGCGTATGTAACGAAGGATCAGAAACTCGA 3308
1368 GGGAGGGCTCCAGTGCAGTTGGCTCCTCAGTGTGCTATGAAAGACCTGAGACAAGG 1427
3309 GAGACGGTATGAGAGCGGAGCTTTGGTCTTGTAGTGTGCTGGGATATGCTGTATCGAT 3368
1428 CAGCTGCTCTGACAGCAGGTGCTTGTCTGAGTGACAAACGGCATCTGCTGTATCGAT 1487
3369 GAGTTCGACAAATGCTGATATGCGCGAGCATGCTTCATGAGGTAAATGGAGCAACAA 3428
1488 GAGTTCGACAAATGATGAAGTAAAGTATGATGATGATGATGATGATGATGATGATGAT 1547
3429 ACCTGATCTGTAGCAGGAGGCTATCATTTGCTGCTGCAACGCTCGGAGCTGTGCTCTT 3488
1548 ACTCTGCTCAATGCAAGGCTGGATCATCTGTGAGTCAATGGCGCAGCTCTGCTCTG 1607
3489 GCATGTGCAAAATCCTAGTGGGTCCGATACAAATGGCGGCTTTCTGTGATGATGATGATG 3548
1608 GCAGCAGCAAAATCCCATTTGAGTCTCAGTGGAAATCTTAAAAAAACAACCATTTGAAA 1667
3549 CAGCTTCTCCAACTCTACTTTCTAGATTTGATTTTAAATTTTAACTTAACTTAACTTAA 3608
1668 CAGCTGCTCATACTTTTATATCAAGGTTTGAATTTGATTTCTCTGCTGCTGCTGCTGCT 1727
3609 GAGCAGCAAAATCCATCGTCTCTGCGCAGGATCTCTGTGGCTTTTAACTATGAAATCTAT 3668
1728 GAGCAAGCCTATGACAGGCTCTGGCTCACCACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787
3669 GAAGTTTCAAGCAGGACGCTTTAGATCTTCAAAACACTTTAACCGGATATATCACTATGCT 3728

Db	1788	GAGCAGCGCAGGAGGAGCTCTCGGACATCGCGGTGCTAAAGGACTACATGTCCTACGCA	1847
Qy	3729	CGTCAGCATGTACATCCTACATTAAGTGTAGGAAGCTCTGAAGATTTGATTAATGGCTAT	3788
Db	1848	CACAGCACCATCATGCCGCGCTAAGTGTAGGAAGCCAGCGCTCTCATCGAGGCTTAT	1907
Qy	3789	GTTGAGATGCGCCAAAGGGCACTTTCTTGGAGCAGTAAAGGTGATAACAGCCACA	3848
Db	1908	GTAGACATAGGAAGATTGGCAGTAGCCGGGAA-----TGGTTTCTGCATAC	1955
Qy	3849	CCTCGGCAACTCGAAAGATGATTCGTATCAGTGAAGCCCTAGCTCCAATGAGATTTCT	3908
Db	1956	CCTCGACAGCTAGAGTCATTAATCCGCTAGCAGAAGCCCATGCTAAAGTAAGATTGCT	2015
Qy	3909	GAAGTGTGAGAAAAGTTGATGCAAGCAGAGCTGTGCGCTTTTAGACGTGCTTTCAG	3968
Db	2016	AACAAAGTTGAAGCCATTGATGTGGAGAGGCGCAAGCGCTCCANNNNNNNGCTCTGAAG	2075
Qy	3969	CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCAAGACTGGAGTG	4028
Db	2076	CAGTCTGCACTGATCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGNNNNN	2135
Qy	4029	TCGSCCAGCGAGCGTATTGTCGCGGCCAACTTGCTAGCTGCTGCGAGAGCTTATAGCA	4088
Db	2136	NN	2195
Qy	4089	GATAAAATTTCCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTCTTCTTGAGGATATC	4148
Db	2196	NN	2255
Qy	4149	CGGAGCCAAAGCAGCTGTGGACGTTAGTTTGCAGGA	4183
Db	2256	NNNNNNNNNGCAATTAATAAGATATGTTTCAAGA	2290

RESULT 12
 CF449139
 LOCUS
 DEFINITION
 CF449139 798 bp mRNA linear EST 04-SEP-2003
 EST685484 normalized cDNA library of onion Allium cepa cDNA clone
 ACABW59, mRNA sequence.
 CF449139
 VERSION
 CF449139.1 GI:34471841
 EST.
 KEYWORDS
 SOURCE
 Allium cepa (onion)
 ORGANISM
 Allium cepa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
 Allium.
 REFERENCE
 1 (bases 1 to 798)
 Havery M.J., Cheung F., Van Aken, S., Utterback, T. and Town, C.D.
 Expressed Sequence Tags from a normalized library of mixed onion
 tissues (Allium cepa)
 Unpublished (2003)
 JOURNAL
 Contact: Havery MJ
 Department of Horticulture
 USDA-ARS and University of Wisconsin
 1575 Linden Drive, Madison, WI 53706, USA
 Tel: 608-262-1830
 Fax: 608-262-4743
 Email: mjhavery@facstaff.wisc.edu
 TIGR sequence name ACABW59PR. For more information:
 http://haverylab.hort.wisc.edu
 Seg primer: CAG GAA ACA GCT ATG ACC.

FEATURES	source
Location/Qualifiers	1..798
/organism="Allium cepa"	
/mol_type="mRNA"	
/cultivar="Red Creole(bulbs), unknown(callus), Ebano & Texas Legend(roots)"	
/db_xref="taxon:4679"	
/clone="ACABW59"	
/tissue_type="Callus, roots, and young bulbs"	
/clone_lib="normalized cDNA library of onion"	

ORIGIN		/notes="Vector: pCMVSPORT6.1-ccdb (Invitrogen); Site_1: EcoRV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA from callus, roots, and young bulbs were combined to synthesize the library. Normalization to enrich for low-copy transcripts was performed by proprietary techniques of Invitrogen."			
Query Match	9.3%;	Score 404.2;	DB 6;	Length 798;	
Best Local Similarity	72.5%;	Pred. No. 9.3e-113;			
Matches 523;	Conservative 0;	Mismatches 198;	Indels 0;	Gaps 0;	
Qy	3376	ACAAATGTCGTGATAATGCCCAGGAGCATGTTCTATGAGGTAAATGGAGCAACAACCGGTAT	3435		
Db	1	ATAAGATGTCGTGATAATGTCGTAGTATGTTACATGAGGTATATGGAAACAACAACGTGAT	60		
Qy	3436	CTGTAGCCAAAGGGGTATCATTTGCTCGCTGAACCTCGGAGCTGTGCTCTTGGCATGTG	3495		
Db	61	CAATTGCAAAAGCTGGTAAATCGCTTCTCTGAATGCACGTACTTCTGTTTTAGCCTGTG	120		
Qy	3496	CAAAATCCTAGTGGGTCCCGATACAAATCGCGCCCTTCTGTGATTGATAACATCCAGCTTC	3555		
Db	121	CAAAATCCAAATGGGTCTGGTTATTAACCCCTCGAATGCTGTGATCGAGAACATTCACCTTC	180		
Qy	3556	CTCCAACTCTACTTTCTAGATTTGATTTAAATTTACTTAATGCTCGCAAAACGACGAGC	3615		
Db	181	CTCCAACTTTACTATCTAGATTTGATTTGATCTACTTGTATCTGGACAAAGCAGACGAAC	240		
Qy	3616	AAAACGATCGTGTCTCGCCAGGACATCTGTGTGCTTTACACTATGAAAACCTATGAAGTTT	3675		
Db	241	AAATGGGATAGACGGTTGGCTTAAGCATATTTTCCCTTGCATTTTGGAAACCCCTGAGAGTG	300		
Qy	3676	CAAAAGCAGGACGCTTAGATCTACAAACACTTACCGGCTATATCACCTATGTCTCGTCAGC	3735		
Db	301	TGGTACAAAGATGCTTGGATCTTCTACGTTACTTTCATAGCTTAGCTATGCAAGAAAGC	360		
Qy	3736	ATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGA	3795		
Db	361	ATATTCAATCCAAAGTTATCCGATGAAGCTGCTGAAGAACTGACTCGTGGCTATGTTGAAA	420		
Qy	3796	TGCGCCAAAAGGGCAACTTTCTGGAGCAGTAAAGGTGATAACAGCCACACCTCGGC	3855		
Db	421	TGAGGAAAAGAGGAAACACCTCTGGTAGCAGTAAAGGGTGAATACAGCCACTGCTAGGC	480		
Qy	3856	AACTCGAAAGTATGATTTCGTATCAGTCAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGG	3915		
Db	481	AAATTGAGATTTGATTTGGCTGAGTGAGACATGGCTCGAATGCGATTTCTCAGAAATGG	540		
Qy	3916	TAGAGAAAGTTGATGACAGCAGAGCTGTGCGCCCTTTTAGACGTGCTTTTGACGAAATCTG	3975		
Db	541	TAGAAGCAGGTGATGTGCAAGAGCGTTTAGGCTACTTGATGTTGCAATGCAGCAGTCTG	600		
Qy	3976	CTACTGATCATGCAACAGGTACGATAGACATGGATCTTTATCAGCACTGGAGTGCGCCA	4035		
Db	601	CAACTGATCATGCCACTGGTAGCATGGATCTTATAATGACCTGGAGTTTCCGCTA	660		
Qy	4036	GCGAGCGTATTTCTGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTTATAGCAGATAAA	4095		
Db	661	GTGAAAGATGAGACGTGAATAATCTAGTTTCAGCAACTCGTGCCTTATATGATGATAAA	720		
Qy	4096	T 4096			
Db	721	T 721			
RESULT 13					
DR917805					
LOCUS	DR917805	709 bp	mRNA	linear	EST 02-AUG-2005
DEFINITION	EST1109344 Aquilegia cDNA library Aquilegia formosa x Aquilegia pubescens cDNA clone COL1X60, mRNA sequence.				
ACCESSION	DR917805				
VERSION	DR917805.1	GI:71687168			
KEYWORDS	EST.				

SOURCE Aquilegia formosa x Aquilegia pubescens
ORGANISM Aquilegia formosa x Aquilegia pubescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Ranunculaceae; Aquilegia.
REFERENCE 1 (bases 1 to 709)
AUTHORS Hodges, S.A., Rensink, W., Buell, C.R., Borevitz, J., Kramer, E.,
Nordborg, M. and Tomkins, J.
TITLE Generation of ESTs from Aquilegia
JOURNAL Unpublished (2005)
COMMENT Other ESTs: EST1109343
Contact: Scott Hodges
Department of Ecology, Evolution and Marine Biology
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
Tel: 805 893 7813
Fax: 805 893 4724
Email: hodges@lifesci.ucsb.edu
Seq primer: M13 Reverse.
FEATURES
source
1..709
/organism="Aquilegia formosa x Aquilegia pubescens"
/mol_type="mRNA"
/db_xref="taxon:338618"
/clone="COLX60"
/tissue_type="mixed shoot and floral apical meristems,
flower buds, leaves and roots"
/lab_host="DH10B T1 (T1 and T5 phage resistance)"
/clone_lib="Aquilegia cDNA library"
/note="Vector: pCMV SPORT6.1; Site 1: EcoRI; Site 2: NotI;
F2, F3, and F4 lines of Aquilegia formosa x A. pubescens
were grown from seed in greenhouses at UC Santa Barbara.
From these plants three sets of tissue were collected: 1)
Small flower buds (<10 mm) and very young inflorescences
(71 & 29% by weight respectively). 2) Medium (7-20 mm) and
large (at or near anthesis) flower buds (65 & 35% by
weight respectively) and 3) Shoot apical meristems. A
fourth set of tissue was collected from plants of A.
formosa. These plants were grown from seed in sand and at
approximately 1 month root tissue and leaf tissue of
various developmental stages were collected (84 & 16% by
weight respectively). Total RNA was extracted from each
set of tissue and pooled in the following proportions:
1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
total RNA, mRNA was extracted and enriched for full-length
messages and then normalized with proprietary methods by
Invitrogen."

Query Match 9.1%; Score 394.4; DB 8; Length 709;
Best Local Similarity 72.3%; Pred. No. 9.6e-110;
Matches 512; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
Qy 3105 TTTCGTTGGAGGCTAAGAAAATTCATCTCGAGCATCTTCCGAGGTGACATCAATGTT 3164
Db 2 TTTCGTTGGAAATGCTTTGAAGCTGCTACAGGTGCTACCTTCGTTGGTGCATCAACAT 61
Qy 3165 TTACTTTGTTGGGACCCCTGGTACAGATAAATCTCAGCTGCTTCAGTATGTGCACAAGATA 3224
Db 62 CTGCTGTCGGGGATCTTCGGGACTAGCAGTCCAGCTGCTCCATATATACACAGCTT 121
Qy 3225 GCTCCTCTGGTAATCTACACTAGTGGGCGAGGAAGTTTCGGCGTTGGGCTGCACAGCGTAT 3284
Db 122 TCCCTCTGGGGAATATATACCAGTGGAGAGGGAGTTCCAGCTGTGGGTTAACTGCGTAT 181
Qy 3285 GTACGAAGGATCCAGAACTCGAGACCGTATTTGGAGAGCGGAGCTTTGGTCTTAGT 3344
Db 182 GTTACTAAGGACCCAGAAACCGGTGAACAGATATTAGAGAGTGGAGCCCTGGTCTGAGC 241
Qy 3345 GATCGTGGATATGCTGTATGATGAGTTCCACAAAATGTCTGATATGTCGCCGAGCATG 3404
Db 242 GATAGAGGCATCTGCTGCAITTGATGAATTTGACAAAATGTCTGAGAATGCAAGGATATG 301

ORIGIN
3405 CTTTCATGAGGTAAATGGAGCAACAAACGGTATCTGTAGCAAAAGGGGGTATCATTTGCTCG 3464
Db TTGCATGAGGTAAATGGAGCAACAAACGGTATCTGTAGCAAAAGGGGGTATCATTTGCTCT 361
Qy CTGAACGCTCGGAGCTGCTCTTCGATGTCGAATCTCAATCTAGTGGGTCGCGATACATCGG 3524
Db CTTAATGCGCAGACATCATGTAAGTCTAGTCTGTGCAAAACCTAGTGGGTCGCGTATATCT 421
Qy CGCCTTTCTGTGATTTGAATCAACATCCAGCTTCTCTCAACTCTACTTTCTAGATTTGATTTA 3584
Db CGTTTATCTGTAATCGACAATATCCACCTTCTCTCAACATTTACTTTCCAGGTTTGAATCT 481
Qy ATTTTACTTAATGCTCGCAAAACCGACAGCAAGCAAAAGATGCTGCTCGCGAGGATCTC 3644
Db ATATACTTAATGCTTGCACAGGCTGATGAACATATGGAATAGGCGCTTCTTGTCTAAACATAT 541
Qy GTGGCTTTTACACTATGAAACTATGAAAGTTTCAAGAGGAGGAGCGCTTAGATCTACAAACA 3704
Db GTGGCACTTCAATTTTGAATCTCTGAGACTGTGCAAGCAAGATGCTTTAGACATACCAACA 601
Qy CTTACCGGTATATCACCTATGCTGCTGATGCTGATGATGATGATGATGATGATGATGATGAT 3764
Db TTGACTGCATATTTGAGTTATGCGCAGAGGCACATTCGCCCAAAATTTGTCAGATGAGCT 661
Qy GCTGAAGATTTGATTAATGCTGCTATGTTGAGATGCGCCAAAGGGCAAC 3812
Db GCAGAAGAAATTTGACTAGAAAGATATGTTGAGATGAGGAGAGAGGAGAAAC 709
RESULT 14
AK089999
LOCUS Mus musculus brain CRL-1443 BC3H1 cDNA, RIKEN full-length enriched
DEFINITION library, clone:G430050N09 product:cmini chromosome maintenance
deficient 4 homolog (S. cerevisiae), full insert sequence.
ACCESSION AK089999.1 GI:26354818
VERSION AK089999.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
PUBMED 10349636
AUTHORS 2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalizaton and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE 3
AUTHORS 3
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, K., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE 4
AUTHORS 4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5

AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
Nature 420, 563-573 (2002)

REFERENCE
6 (bases 1 to 3285)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayaashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
Direct Submission

JOURNAL
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
Location/Qualifiers
1..3285
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H"
/db_xref="taxon:10090"
/db_xref="FANTOM_DB:G430050N09"
/clone="G430050N09"
/cell_line="CRL-1443 BC3H1"
/tissue_type="brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
125..2713
/note="unnamed protein product; mini chromosome maintenance deficient 4 homolog (S. cerevisiae) (MGDI:103199, GB|NM_008565, evidence: BLASTN, 100%, match=3221)
putative"
/codon_start=1
/protein_id="BAC41036.1"
/db_xref="GI:26354819"
/translation="MESPATPSRRSRGRVTPQSLRSESRSPNRRGDSST
GELLPMPTSGADLQSPAPNALFSSPPQMHSLAIPDLFDVSSPLTIGTPSRVGGTP
RSGVTRPVRQDLPAGARKGLQVDSGAAAEIDVPSQGLQKLVIGTVDNVAT
CKENFORLECFDPLQAEBEENVGIDITQPLYNQQLGEINIITGEPLVNCBHIFS
KNLYRLIISYPQEVIPTEDMVNEIIFEDRVPDSILEHQIQRPFNALTKSMRNLP
DIDLLITISGMVIRTSQILPEMQEAFQVQVCAHTTFRVLDGRGIAEPCSCVCHHTH
SWALIHNRFFSDQMKIQESPEMPAGTPTHTIVLFAHNLVDVKVQPDVRVNTGI
YRAPVTRVPRVSNVSVYKHIDIVHYRKTDAKRLHGLDDEABQKLFSEKRVLLKE
LSRPDIYERLASALAPSIVYEHEDIKKGILLQLFGGRTDFSHGTGRFRAEINILLC
GDPGTSKQLQVYVNLVPRGOYVTSKGSAGVLTAYVMKDPETROLVLTQALVLS
NGICCDKDKMNESTRSVLHVEWEOQLSIAGAGIICOLNARTSVLAANPIESOWN
PKKTIENIQPLHLSRFDPLIFLMDLPQDEANDRLAHLHLVSLYQSEOEVEEFLD
MAYLVKDYIAYHSTIEMPRSEEAQALIEAYVNNRKGISGRGMVSYPRQLESILBLA
EHAQVFNFSNKAIVDVEEAKRLHREALKQSDATDPRTGMVDIISILITGMSATSKRKE
ELAEALRLKILSKGTPALKYQQLFEDIRQSDTAITKMFEEALRALADDDFLTVTG
KTVRLI"
3268..3273
/note="putative"
3285

polyA_signal
polyA_site

FEATURES
source
CDS

ORIGIN		/note="putative"	
Query Match	Score 392.8; DB 4; Length 3285;		
Best Local Similarity	53.3%; Pred. No. 5.5e-109;		
Matches 1021; Conservative	0; Mismatches 812; Indels 81; Gaps 6;		
QY	2353 CAACGTTTTCAGAGCATATTGAGGCGAGACCTTTTCATCTCAAGCATCGGTGCATGC 2412		
Db	876 CCATCTTGAACATCAGATTCAGTCAGACCTTTTAATGCGTTTGAAGACAAAGAGTATGA 935		
QY	2413 GTGAATCAACCTTCAGATATAGACAAATTTGTTTCTGTTTAAAGGAATGGTTTCGCGT 2472		
Db	936 GAACTTGAATCCAGAAGACATTTGATCAGCTCATCAACCATCAGTGCATGCTCATCAGAA 995		
QY	2473 GCAGTTCTATCATACCTGAAATTAAGGGGGCCCTTTTCAATGTTTAGTGTGTGTCAC 2532		
Db	996 CATCAAGCTGATTCGGAGATCAGAGGCGCTTTTTCATGCCAAAGTCTGTGCCACA 1055		
QY	2533 CGCTCCGCTAGTTACAGTTTGTAAAGGGCGGTTTGAAGGAGCCAAACAAGGTGTGAAAGC 2592		
Db	1056 CCACCGGGTGGAGA---TAGATCGAGGAGCAAGATTTGCTGAGCCCTGCAGTTGTGTGCACT 1112		
QY	2593 CAGAAATGTGAGCAGCAAGGAATGCTATGCTCTTATTACAAATCGATGCACCTTTTCAATA 2652		
Db	1113 GCCACACTACCCACAGC-----ATGGCACTGATCCCAACCCGATCATTTCTTCTGACA 1166		
QY	2653 AGCAGATAGTGGCTTTTCAAGAACTCCAGATGCCATTCCTGAGGAGAGACTCCACACA 2712		
Db	1167 AGCAAAATGATCAAACTTCAAGAGTCTCTGAGACATGCTGCTGGGCGAGACACTCACA 1226		
QY	2713 CAGTCAGCATGTGTTTATACAAACATATGTTTGAATGCTGGAAGCCTGGAGATGTTATTG 2772		
Db	1227 CTATTGTCTTTTGGCCCAAGTACGACCTTTGTTCAAGGTTCAACCCAGGGGACAGAGTGA 1286		
QY	2773 AGGTAAACAGGAGTTTTCAAGGCCATGAGCTTTCAGTTGGTTCGAAATCAACGAACATTAC 2832		
Db	1287 ACGTACAGGCATATATCGAGCAGTACCAATTCAGATTAATCCAGAGTGAAGCAACGTGA 1346		
QY	2833 GAGCATTTGATAAGACCTTACATCGATTGCGTGCAGCTCAAGAAGTCTGCACAGGGGTGCAC 2892		
Db	1347 AGTCTGTCTATAAAACCCACATTTGATGTCATTCATTATTCGGAACCGATGCAAAACGTC 1406		
QY	2893 TGCAAACTGAAGATCCTATGAGAGATGGATAAGGAGAATGATATGATGCTGGGTATCATG 2952		
Db	1407 TGCA-----TGSCCTTGATG 1421		
QY	2953 AAAGTGATATCTCAGAACTGCTAATGAAGCAAGATTCAAAACCTTAAAGAGCTGTCCA 3012		
Db	1422 AAGAAGCAGAACAGAAACCTTTTTCAGAGAAAACGTTGAAATTTGCTTAAAGGAACCTTTCCA 1481		
QY	3013 AGCTCCCGGCATTTATGATAGACTTTCAAGGTCTGCTGGCTCCAAAGCATTTGGGAGCTTG 3072		
Db	1482 GGAAGCCAGATATTTATGAGCGGCTTGCTTTCAGCTTGGCTCCAGCATTTATGAACATG 1541		
QY	3073 AAGATATTAAAAAGGCTCTCTTTTTCAGCTCTCTTTGGTG-----GGAAGGCTAAAGAAA 3126		
Db	1542 AAGATATCAAAAAGGGAATCTTACTTTCAGCTCTCTTTGGTGGAAACAGGAAGATTTTCAGTC 1601		
QY	3127 TTCCATCTGAGAGCATCTTTCCAGAGTGAATCAATGTTTATCTTGTGGGAGCCCTGGTA 3186		
Db	1602 AACTGGGAGGGGTAATTTCCGCTGCTGAGATCAACATCTCTTGTGTGGGAGCCCTGGCA 1661		
QY	3187 CCAGTAAATCTCAGCTGCTTTCAGTATGTCACAAGATAGTCTCTCGTGGAACTTACACTA 3246		
Db	1662 CCAGCAAGTCCAGCTGCTACAGTATGTGTACAACTGTGTCCTCCAGAGGCCAGTACAGT 1721		
QY	3247 GTGGGCGAGGAAGTTCGCGGCTTGGGCTGACAGCTATGTAACGAAGATTCAGAAATC 3306		
Db	1722 CTGGAAGAGGCTCCAGTCGCGCTCCGCTCACCCTATGTGATGAAAGACCTTGAGACCA 1781		
QY	3307 GAGAGACGGTATTGGAGAGCGGAGCTTTGTTCTTATGATGCTGGGATATGCTGTATCG 3366		

Db 1782 GCGAGCTTGTCTCCAGACAGAGTGCCTCGTCTGATGACAAATGGGATATGCTGTCATCG 1841

QY 3367 ATGAGTTCGACAAAATGCTGTGTAATGCCCGAGCATGCTTCATGAGGTAATCGAGCAAC 3426

Db 1842 ATAAGTTTGACAAAATGATGAAGCACAAGTCTGTGCTGATGAGTTCATGGAACAGC 1901

QY 3427 AAACGGTATCTAGCCAAAGGGGTATCATTCGCTCGTGAACGGTCGGAGCTCTGTGCC 3486

Db 1902 AGACTCTGTGTCATTCGAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCC 1961

QY 3487 TTGCATGTCAAAATCTAGTCCGATACAACTGACGCGCTTCTTGATGTGATTAACA 3546

Db 1962 TGGCAGCAGCAATCTTATGATCTCAGTGGATCTTAAACCAACCATGTGAATA 2021

QY 3547 TCCAGTCTCTCCAACTCTACTTCTTAGATTTGATTTAATTTACTTTAATGCTCGCAAC 3606

Db 2022 TCCAACTACCGCACATTTGTCAGAGTTTGATCTCATTTCTCTCATGCTAGACCTC 2081

QY 3607 CAGACGAGCAAAACGATCTGCTGTCGCCAGGATCTCGTGGCTTTACACTATGAAAAC 3666

Db 2082 AGGATGAGGCATATGACCGCGCTAGCTTCATCAGCTGTTCAITGTTACTACCAAGTG 2141

QY 3667 ATGAAGTTTCAAGCAGGAGGCTTAGATCTCAACACTTACCGGTATATCACCTATG 3726

Db 2142 AGGAGAGTGGAGGAGGATTTCTGAGCATGGCCGTCGGAAGACTACATTCGATATG 2201

QY 3727 CTCGTGAGCATATCATCTACATTAAGTGAAGAGCTCTCAAGATTTGATTAATGGCT 3786

Db 2202 CCCATAGTACCATATGCCCCGACTGAGTGAGGAGCCAGCGCTCTCATTTAGGCTT 2261

QY 3787 ATGTTGAGATGCGCAAAAGGGCACTTTCCTGGAAGCAGTAAAGGTGATAACAGCA 3846

Db 2262 ATGTAACAATGAGGAAGAT-----TGGGAGTAGCGGGGATGGTTCTTCGCTT 2309

QY 3847 CACCTCGGCAATCCGAAGTATGATCGTATCAGTGAAGCCCTAGCTCGAATGAGTTT 3906

Db 2310 ACCCTCGACAGCTAGAGTCATTAATTCGCTTAGCAGAGCCCATGCTAAAGTAAGATT 2369

QY 3907 CTGAAGTGTAGAGAAAGTTGATGACAGAGAGCTGTGGCCCTTTTAGACGTCGCTTGC 3966

Db 2370 CAAACAAGTTGAGCAATGATGTGAAGAGCAAGAGCTCCACGGGAGGCTCTGA 2429

QY 3967 AGCAATCTGCTACTGATCATGCAACAGGTAACGATGACATGATCTTATCAGACTGGAG 4026

Db 2430 AGCAGCTGCAACTGACCTCGTACTGCGATGGTGTGATATTTCTTCTTACTACAGGAA 2489

QY 4027 TGTCCGCGACGAGGCTATTCGTGCGGCCAATTCGTAGCTGCTCTGCGAGAGCTTTAG 4086

Db 2490 TGAGTGCCACTTCTCTGAACGGAAGAGAAATTAGCTGAAGCATTTGAGAAAATTTT 2549

QY 4087 CAGATAAAATTTTCACTGCGAGCTCTCTGCTTGAAGACCAAGTCAAGCTTCTTGAGGATA 4146

Db 2550 TATCTTAA-----GGTAAACACAGAGCTTAAAGTACCAAGCTGTTTGAGGATA 2600

QY 4147 TCCGAGCCAAAGCAGTGTGAGCTTGTGAGGATATTAATAATGCTCTGGGTAGCC 4206

Db 2601 TTCCGGGAGCAGCTGACACAGCAATTAACCAAGACATGTTTGAAGAGCCCTCGAGCTT 2660

QY 4207 TCCAAGGAGAGGCTTTCTTACTGTCATGTTGATGATGATGATGATGATGATGATGATG 4260

Db 2661 TGGCTGATGATGATTTCTTAAACAGTGAAGAGTGTGCGGCTGCTGTGAG 2714

RESULT 15

AK088796

LOCUS

DEFINITION AK088796 3301 bp mRNA linear HTC 03-APR-2004
Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
full-length enriched library, clone:B430026B21 product:mini
chromosome maintenance deficient 4 homolog (S. cerevisiae), full
insert sequence.

ACCESSION

AK088796

VERSION

AK088796.1 GI:26353895

KEYWORDS

HTC; CAP trapper.

SOURCE
ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/WRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>.

FEATURES	Location/Qualifiers	Source
	1. .3301 /organism="Mus musculus" /mol_type="mRNA" /strain="NOD" /db_xref="FANTOM DB:E430026B21" /db_xref="taxon:10090" /clone="E430026B21" /cell_type="thymic cells" /tissue_type="thymus" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days neonate" 141. - 2729	
CDS	/note="unnamed protein product; mini chromosome maintenance deficient 4 homolog (S. cerevisiae) (MGD MG1:103199, GB NW_008565, evidence: BLASTN, 100%, match=3221) putative" /codon_start=1 /protein_id="BAC40578.1" /db_xref="GI:26353896" /translation="MSPASTPSRRSRREVRVPTQSLSESRSSPNRRRGDSST GELLPPGTQRPDQSPNALFSSPPQKHSLAIPLDFDVSPILTYGTPESSRVGTP RSGVPTQVRQDLGSARKGVLDQSGAAAEIVPSEQLSGKLVIWGDVNVAT CKENFORFLQDTPLAKEENYIDITQPLYMQLGEINITGEEPLFNVNCSHKISFS KNIYROLISYQSVIPTFDMVANEIIFDRYPSVILEHQIQRVFNALTKSMRNLPSE DIQDILTISGMVIRTSLQIPEMOEAFPOQVCAHTRVEIDRGRIAPGPCSVCHCHTH SMALLHNRSFSPNKQMLQESPEDMPAGOTPHIIVLFAHNDLVDVQPGDVRNVGTI YRAVIRNPRVSNKSVYKTHIDVIHYKRTDKRLHGLDEAEOKLFSEKRVKLKE LSRKPDITLAEALAPSIYEHEDIKGIILLQJLFGGRKDFPSHTGRKPRASINLLC GDPTSKSOLLQVNLVNPVPRGVSGSAVLGTAYVMKDPETROLVLOTGALVUSD NGICIDIEPKOMNESVRLHVEWEOQTLSIAKAGII COLNARTSYAAANPIESOWN PKKTTIENIOLPHTLLSRPLIIFMLDPOEAYDRRLAHLVSLYYQSEEOVEEFELD MAVLKDIYAHNSTMIPSEEAQALIEAYVMRKIGSRGMVSAVPRQLESILRLA EAHAKRFSKNVRKALDVEERKRLHREALKQSATDPRTGIVDISILITGMSATSRKKE ELAELARKLILSKGKTPALKIQQJLFEIDIRGQSDTAITKMFELRALADDDFLVTG KTVRLI" polyA_signal 3284. .3289 /note="putative" polyA_site 3301 /note="putative"	
ORIGIN	Query Match 9.0%; Score 392.8; DB 4; Length 3301; Best Local Similarity 53.3%; Pred. No. 5.5e-109; Matches 1021; Conservative 0; Mismatches 812; Indels 81; Gaps 6; 2353 CAACGTTTGAGNAGCATATTGAGGCCAGACCTTTCAATCTCAAGAGCATCGGTGCACATGC 2412 892 CCATCTTAGAACATCAGATTCAAGTCAGACCTTTTAATGCGTTCGAAGACAAGAGTATGA 951 2413 GTGAACCTCAACCTTTCAGATATAGACAAATTGGTTTCTGTGTTAAAGGAATGGTTATCCGGT 2472 952 GAAACTTGNATCCAGNAGACATTGATCAGCTCATCAACCATCAGTGGCATGTGTCATGAAA 1011 2473 GCAGTTCATCATACCTGAAATTAAGGGGCCCTTCTTCAATGTTTAGTGTTGGTGCACT 2532 1012 CATCAGCTGATCCGGAGATCGAGAGGGCCCTTTTCCATGTCGCAAGTCGTGCGCCACA 1071 2533 CGCCTCCGCTAGTTACAGTTGTTTAAAGGCGGGTTGAGAGCCCAACAGAGGTGTGAAAAGC 2592 1072 CCACCCGGGTGGAGA---TAGATCAGAGGAGAAATTTGCTGAGCCCTGCAGTTGTGTGCAT 1128 2593 CAGAAATGTGCAGCAGGAATGCTATGTCTCTTTATTCACAAATCGATGCACTTTTGGCAATA 2652 1129 GCCACACTACCCACAGC-----ATGGCACTGATCCACAAACCGCATCTTCTCTGACA 1182 2653 AGCAGATAGTGGCTCTTCAAGAAACTCCAGATGCCATTCCTCCTGAAGGAGAGACTCCACACA 2712 1183 AGCAAAATGATCAAACTTCAAGAGTCTCCCTGAAGACATGCTCTCTGGCAGACACCTGACA 1242 2713 CAGTCAGCATGTGTTTATACACACATATCATGTTTGATGCTGTGTGAAGCGCTGGAGATCGTATTG 2772	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:03:05 ; Search time 191 Seconds
(without alignments)
2072.672 Million cell updates/sec

Title: US-10-768-511-8

Perfect score: 4609

Sequence: 1 MENDNALDIGAVSPYPSQS.....LGSLOGEGFLTVHGDIVKRV 901

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4609	100.0	901	5	AAG66003 P. patens
2	4609	100.0	901	7	AAG39980 Physcomit
3	2550	55.3	881	8	ADY09589 Plant ful
4	1883	40.9	1023	6	ABJ26500 Aspergill
5	1860	40.4	874	6	ABJ25900 Aspergill
6	1835.5	39.8	989	7	ADN21148 Bacterial
7	1835	39.8	960	8	ADN21148 Bacterial
8	1833	39.8	933	8	ADN21148 Bacterial
9	1817.5	39.4	923	7	ADN18710 Human dis
10	1817.5	39.4	923	8	ADN18710 Human dis
11	1817.5	39.4	924	8	ADN06294 Novel bro
12	1814.5	39.4	863	8	ADN69821 Human min
13	1810.5	39.3	863	4	AAB95203 Human pro
14	1810.5	39.3	863	8	ABM80987 Tumour-as
15	1754	38.1	910	5	ABP73643 Candida a
16	1732.5	37.6	918	8	ADP83536 Breast sp
17	1686.5	36.6	866	4	ABM58299 Drosophil
18	1670.5	36.2	456	8	ADY13148 Plant ful
19	1647	35.7	820	8	ABM82800 Human dia
20	1500	32.5	712	8	ADQ19271 Human sof
21	1375	29.8	724	3	AAB56475 Human pro
22	1199.5	26.0	656	8	ADP83535 Breast sp
23	1071	23.2	682	4	AAB62036 P. furios
24	1020.5	22.1	666	8	ADS43108 Bacterial

25	1016.5	22.1	760	8	ADS44134	Bacterial
26	1003	21.8	814	6	ABJ26042	Aspergill
27	997.5	21.6	287	8	ADX68835	Plant ful
28	985.5	21.4	718	6	ABJ25741	Aspergill
29	983.5	21.3	713	6	ABJ26341	Aspergill
30	983.5	21.3	809	5	ABP73372	Candida a
31	976.5	21.2	720	4	ABB61153	Drosophil
32	973.5	21.1	632	6	ABJ25442	Aspergill
33	973	21.1	699	8	ADS44402	Bacterial
34	972	21.1	743	8	ADX95960	Plant ful
35	970	21.0	766	3	AAB43937	Human can
36	963.5	20.9	720	5	AAO22751	Maize pro
37	963	20.9	775	6	ABR53639	Protein s
38	963	20.9	775	7	ADK63268	Disease t
39	963	20.9	775	8	ADN19147	Bacterial
40	961.5	20.9	719	8	ADO19649	Human PRO
41	961.5	20.9	719	8	ADO19790	Human PRO
42	961.5	20.9	719	9	ADU06531	Novel bro
43	961.5	20.9	719	9	ADY16524	PRO polyp
44	961.5	20.9	881	8	ADN22948	Bacterial
45	960.5	20.8	719	8	ADQ09276	Human MCM

ALIGNMENTS

RESULT 1

AAAG66003	ID	AAAG66003	standard; protein; 901 AA.
XX	AC	AAAG66003;	
XX	DT	27-FEB-2002	(first entry)
XX	DE	P. patens	cell cycle protein 2 (CC-2).
XX	KW	Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;	
XX	KW	CC-3; environmental stress.	
XX	OS	Physcomitrella patens.	
XX	PN	WO200177354-A2.	
XX	PD	18-OCT-2001.	
XX	PF	06-APR-2001; 2001WO-US011294.	
XX	PR	07-APR-2000; 2000US-0196001P.	
XX	PA	(BADI) BASF PLANT SCI GMBH.	
XX	PI	Costa E SilvaO, Bohnert HJ, Van Thielens N, Chen R;	
XX	PI	Sarria-Willan R;	
XX	DR	WPI; 2002-049151/06.	
XX	DR	N-PSDB; AAI67613.	
XX	PT	Novel Cell Cycle Stress-Related Protein useful for increasing tolerance	
XX	PT	to environmental stress, is selected from Cell Cycle Proteins 1-3, or	
XX	PT	their orthologs.	
XX	PS	Claim 3; Fig 3B; 90pp; English.	
XX	CC	The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),	
XX	CC	isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1	
XX	CC	protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSRP and	
XX	CC	encoding nucleic acids are useful for increasing tolerance to	
XX	CC	environmental stress selected from monocot and dicot selected from maize, rice,	
XX	CC	transgenic plants including monocot and dicot selected from maize, wheat,	
XX	CC	rye oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,	
XX	CC	canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,	
XX	CC	tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,	
XX	CC	tea, Salix species, oil palm, coconut, perennial grass and forage crops.	

Db 352 LSSGPAVSSDALSETGRMPQNEIPDLVGEVETKAFKVLPPGLDSTVNMRLDLPADMDKL 411
Qy 298 VSVKGMVIRCSIIPEIKGAPFKCLVCGHSPPLVTVWGRVVEPTCEKPECAARNAMSL 357
Db 412 VSIKGLVIRTTPIIPDMKEAFRCQVNHGVQ-VDIDRGKIAEPTCEPRPVCKERNMQL 470
Qy 358 IHNRCCTFANKQIVRLQETPDALPEGETPHTVSMCLYNTVMVDVAVKPGDRIEVTGVFKAMAV 417
Db 471 IHNRCVFAKQIVRLQETPDSIPDQTPHSVSLCVYDELVDVCKAGDRVETGVIFRCNPV 530
Qy 418 RVGNPQRTLRALYKYIYDCVHVKKSDRGRLQETDPMEMDKENDMYAGHESDTSEANEA 477
Db 531 RVNPRQRTQKSLFTYIDVHLVQKIDRKGLGDVSTIEQELSEQAAGDAEQTRRLTAEE 590
Qy 478 KIQKELSKLGIYDRLSRSLAPSWELEDIKGLLQCLFGGKAKKIPSGAS--FRGDI 535
Db 591 --EKIKRTATRPDLVELLSRSLAPSIYEMDDVKKGILLQFGTKNTFKGKNPRYRGI 648
Qy 536 NVLLVGDPGTSKQLLYVHKIAPRGITYSGSSAVGLTAYTKDPETRETVLSEGLV 595
Db 649 NILCGDPSTSKQLLYVHKIAPRGITYSGSSAVGLTAYTRDPETRQMWLESGALV 708
Qy 596 LSDRGICCIDPDKMSDNARSMLHEVMEQOTVSVAKGIIASLNARTSVLACANPSGSRV 655
Db 709 LSDGGICCIDPDKMNESTRSVLHEVMEQOTVSIAKAGIITTLNARTSLASANPIGSRV 768
Qy 656 NARLSVIDNIQLPPTLLSRFDLIYMLDKPDQNDRLARHLVALHYEN--YEVSKODAL 713
Db 769 NNPLVPQNIIDPPTLLSRFDLIYVLDVRDQEDRLAKHLVNMYLEDRPEHAHQEIL 828
Qy 714 DIQTTAYITYARQHVHPTLSDEAAEDLINGVEMKQGNFPGSSKKVITATPRQLESMI 773
Db 829 PIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMKLGDDIRSSDRITATTROLESMI 888
Qy 774 RISEALARMFSEVKEVDAARVLLDVALQASATHTATGIDMDLITTVGSASERIRR 833
Db 889 RLSEAHARMRLSPVTTADVEAVRLIRSAIKQAATDSRTGLIDMSLLTEGTSASERRSR 948
Qy 834 ANLLAALRELIADKISPGSSGLKTSQLEEDIRSQSSVDVSLQDIKNALGSLQGEGLTV 893
Db 949 EALKRALLSVVDLDCGGGAA--RWAEEVFRILSENSSIEVDGAQFADAVRALEAGAVSV 1006
Qy 894 HGDIVKR 900
Db 1007 VEGARR 1013

RESULT 5
ID ABJ25900
XX ABJ25900 standard; protein; 874 AA.
AC ABJ25900;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #558.
XX
XW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
FN W0200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.

PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
XX Sequence 874 AA;
Qy
Query Match 40.4%; Score 1860; DB 6; Length 874;
Best Local Similarity 45.5%; Pred. No. 6.3e-146;
Matches 413; Conservative 146; Mismatches 244; Indels 104; Gaps 18;
Qy 51 TPTSAVRRGRETDSARRRRRSRLGNSVSPYDAGTGTGTPVATPVATPVGTPM 110
Db 5 SPTRAQNRQRRSDIP-----SSSGLFVSSRP----- 32
Qy 111 GTPSPHRTGPQYKQSELGSGQ--KPLHRRRSQSREPQHRSPSPSADGRPSEGA--- 165
Db 33 -SIESNRV---SRRSDLHSGGFLSPNRRRVFVDANG-----MPATDGRSPDATFS 82
Qy 166 --EPD----DTLGGEYA-YVWGTNVNIPDLVLAIRFLHN-----YRSSADHLNSKYI 214
Db 83 NIHPDTSAEALGGSTRVINGTINISIQDSMSAFKNFLYFQTKYRLWAEGATEDETRIM 142
Qy 215 EETVERER-----DTLNIDMSDI--YDHDPDLVYAKIVYPLDIIPLLDTEQC 259
Db 143 GDSABEREYISMLSTMRLQGLVTSINLDKAKLKPSTKLHWHQLHAYPOBIIPLMDQTVK 202
Qy 260 EVATSL-----LPTFE-KHTEARPFNLKASVHMEELNPSDIDLK 297
Db 203 DVNVELAIKEMERLRAQNQRNQHNRGLSSVETKAFKVLPPGLDSTVNMRLDLPADMDKL 262
Qy 298 VSVKGMVIRCSIIPEIKGAPFKCLVCGHSPPLVTVWGRVVEPTCEKPECAARNAMSL 357
Db 263 VSIKGLVIRTTPIIPDMKEAFRCQVNHGVQ-VDIDRGKIAEPTCEPRPVCKERNMQL 321
Qy 358 IHNRCCTFANKQIVRLQETPDALPEGETPHTVSMCLYNTVMVDVAVKPGDRIEVTGVFKAMAV 417

```
Db 322 IHNRCVFADKQVVKLQETPDSIPDGTQPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 381
Qy 418 RVGNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMENDKENDMTYAGHESDTSANEA 477
Db 382 RVNPRQTKSLFKTYIDVLHVQKIDRKKLGDVSTIEQELSEQAAGDAEQTRRLTABEE 441
Qy 478 KIQKLKSLKPGIYDRLSRLAPSWELEDIKKGLCOLFGGKAKIPSGAS--FRGDI 535
Db 442 --EKIRATRPDLYELLSRLAPSIYEMDDVKKGLLQLFGGTWTKTQKGNPRYRDI 499
Qy 536 NVLLVGDPTSKSOLLQYVHKIAPRGYITSGRGSAGVLTAYVTXDPETRETIVLSGALV 595
Db 500 NILLCGDPSTKSQLLRVYVHKIAPRGVYITSGKSSAGVLTAYVTRDPETRQWVLSGALV 559
Qy 596 LSDRGICCIDPDKMSDNARMLHEVMEQQTVSVAKGGLIASLNARTSVLACANSGSRY 655
Db 560 LSDGGICCIDPDKMNESTRSVLHEVMEQQTYSIAKAGIITTLNARTSILASANPIGSRY 619
Qy 656 NARLSVIDNIQPLPTLLSRFDLIYMLDKPDRDRRLARHLVALHYEN--YEVSKODAL 713
Db 620 NPNLPVQNIIDUPFLLSRFDLIYVLDKVDQEDRRRLAKHLVNMWYLEDREPHAAEQEIL 679
Qy 714 DIQTLTAYITAYARQHVHPTLSDEAAEDLINGVEMRQKGNFPGSSKKVITATPROLESMI 773
Db 680 PIEFLTAYITAYAKTVHPVLTPAAGKALSDAVVMRKLGGDIRSSDRITATTRQLESMI 739
Qy 774 RISEALARMFSEVVEKVDAAVALLDVALQOASATDHATGYIDMDLITTVGSASERIRR 833
Db 740 RLSEAHARMRLSPVTDADVEAVRLIRSAIKQAATDRTGLIDMSLLTEGTSASERRSR 799
Qy 834 ANLLAALRELIADKISPGSSSGLKTSQLEDIRSSQSVDSVLSQIDKNALGSIQGGFLTV 893
Db 800 EALKKALLSVDDLCGGGAA--RWAEPRIILSENSIEVDGAQFADAVRALEAEGAVSV 857
Qy 894 HGDIVKR 900
Db 858 VEGGAR 864

RESULT 6
ID ADB70288
XX ADB70288 standard; protein; 989 AA.
XX AC ADB70288;
XX DT 04-DEC-2003 (first entry)
XX DE C. neoformans amino acid sequence SEQ ID NO:3332.
XX KW fungicide; gene therapy; infection.
XX OS Cryptococcus neoformans.
XX PN WO2003052076-A2.
XX PD 26-JUN-2003.
XX PF 17-DEC-2002; 2002WO-US040225.
XX PR 17-DEC-2001; 2001US-0341261P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Zamudio C, Eroshkin AM;
XX DR WPI; 2003-533017/50.
XX DR N-PSDB; ADB69205.
XX PT New nucleic acid, useful for preparing a composition for treating an
XX PT infection caused by Cryptococcus neoformans.
XX PS Claim 9; SEQ ID NO 3332; 136pp; English.
XX
```

The invention relates to a novel purified or isolated Cryptococcus neoformans nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by Cryptococcus neoformans. The present sequence represents a C. neoformans sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 989 AA;

```
Query Match 39.8%; Score 1835.5; DB 7; Length 989;
Best Local Similarity 43.8%; Pred. No. 8.7e-144;
Matches 409; Conservative 147; Mismatches 246; Indels 131; Gaps 20;

Qy 47 AVRETPTSAVR-----RGRETDARRRRSRSLGNSVYSSP-----YD 86
Db 4 ALLQSDSAMKHCSSLLGSSGGSTPRARRGDTHSSPSPSLARRQAQVPADSLSLR 63
Qy 87 AGTPTPTPTVATPVYATPVGTPTMGTPSFHRTGTPQYKORSBELGSGCKPLHRRRRRSQSRP 146
Db 64 ASSPG-----NGTDSF--GTPRV-----FGSAAPTLSAVAQSQL--- 95
Qy 147 CHRSPSPSPADGRPSSEAPDDTLGGEYAVVGTNVNIPDVLRAIRPL----- 196
Db 96 -----GADG-----DDIDGMVKPIWGTITISLQESMNLFRDPLRGFKPKYRAV 137
Qy 197 HNYRSSADLNS-----KYIQIIEETVE-----REEDTLNIDMSDIYD 234
Db 138 YNASQSRHAESGGVAPPMTLYDNLSEAEVPLYEYTLNRLTGTNMLDLNALLA 197
Qy 235 HDP--DLVYRYVPLDIIPLLDTECEQEVAT-----SLPTTFKHI-E 274
Db 198 YRPTKLYQQLVNPQEVIPIMDQVLRDVMIELGHEELEKAKTFAEGLNSQDVESRVYK 257
Qy 275 ARPNNLKASVMRELNPSSDIDKLVSKGMVIRCSIIPEIKGAFKCLVCGHSPPLVTVV 334
Db 258 VRPFGGKTVNMRDLNPGDTDKLVTKGLVIRATPVIPTMTTAFPRCLVCQHTVQ-ADID 316
Qy 335 KGRVEEPTRCPECAARNAMSLJHNRCTFANKQIVRLQETPDALPEGETPHTVTSMCLYN 394
Db 317 RGRISEPERCDVCGGTGWSLJHNRSEFTDKQVIRLQETPDVDPGQTPTHTVSLCVD 376
Qy 395 TMVDVAVKPGDRIEVTGPFKAMAVRGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPME 454
Db 377 ELVDLVKPGDRIIVTIGIPRSIPVRVNPQRSIKSLYKTYLDVWVVKRTNTARMGF-DPST 435
Qy 455 MDKENDMYAGHESDTSSEANEAKIQKLKELSKLPGIYDRLSRLAPSWELEDIKKGLL 514
Db 436 RAGESKP-PGDLGSPVRSAAAEME-QRIEILSNHPDLYNIIASSLAPSIELEDYKKGIL 493
Qy 515 COLFGGKAKKIPSGAS-----FRGDI NVLLVGDPTSKSOLLQYVHKIAPRGYITSGRGS 569
Db 494 LQLFGGTTKSIARGGGGGGPRYRGDI NVLMWGDPTSKSOLLQYVHKIAPRGYITSGRGS 553
Qy 570 SAVGLTAYVTXDPETRETIVLSGALVLSDRGICCIDPDKMSDNARMLHEVMEQQTVS 629
Db 554 SAVGLTAYVTRDPDSKQLVLSGALVLSGGVCCIDEFDKMSDATRSVLHEVMEQQTVSI 613
Qy 630 AKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQPLPTLLSRFDLIYMLDKPDEQN 689
Db 614 AKAGIITTLNARTSILAAANPINSRYDPNLPIPANIDLPPTLISRFDLVLDKVDVYN 673
Qy 690 DRRLARHLVALHYENYEVSKOD-ALDLOTLTAYITAYARQHVHPTLSDEAAEDLINGYVEM 748
Db 674 DRKLAKHLVGLYLSQVDEQPADNIIPQLTILSYIYIYARSKHPVUTEGSEALVQAYVEM 733
Qy 749 RQKGNFPGSSKKVITATPROLESMTIRISEALARMFSEVVEKVDAAEAVRLDLVALQOSA 808
Db 734 RKAGMDSRTQEKRTTATTTRQLESMTIRISEALARMFSEVVEKVDAAEAVRLDLVALQOSA 793
```


Db 861 TRDDVREAVRLIKSALKAATD-SQGRIDMSLLTGTSAARQKADKDAVIRLLDEMT 919

QY 849 SPGSSGLKTSOLLEDIRSQSSVDVSLQDINKNALGSLQEG 889

Db 920 SGGQV--VRYSEVARRLGGAGVQVEPAEFAEVRMALEMEG 958

RESULT 8

ADS44081

ID ADS44081 standard; protein; 933 AA.

XX ADS44081;

XX

DT 02-DEC-2004 (first entry)

XX

DE Bacterial polypeptide #22511.

XX

KW Recombinant DNA construct; transformed plant; improved plant property;

KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

KW pathogen tolerance; pest tolerance; plant disease resistance;

KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;

KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polypeptide.

XX

OS Bacteria.

XX

PN US2003233675-A1.

XX

PD 18-DEC-2003.

XX

PF 20-FEB-2003; 2003US-00369493.

XX

PR 21-FEB-2002; 2002US-0360039P.

XX

PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

XX

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX

DR WPI; 2004-061375/06.

XX

XX

PT New recombinant DNA construct comprising a promoter positioned to provide

PT for expression of a polynucleotide encoding a polypeptide from a

PT microbial source, useful for producing plants with improved properties.

XX

PS Claim 1; SEQ ID NO 22511; 122pp; English.

XX

CC The invention relates to a recombinant DNA construct comprising a

CC promoter functional in a plant cell, where the promoter is positioned to

CC provide for expression of a polynucleotide encoding a polypeptide from a

CC microbial source. The invention also relates to a transformed plant

CC comprising the recombinant DNA construct and a method of producing a

CC transformed plant having an improved property. The plant is a crop plant

CC such as maize or soybean. The method of producing a transformed plant

CC having an improved property comprises transforming a plant with the

CC recombinant DNA construct and growing the transformed plant, where the

CC polynucleotide or polypeptide is useful for improving plant properties.

CC The recombinant DNA construct is useful for producing plants with

CC improved plant properties, e.g. improved cold, heat or drought tolerance,

CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,

CC increased resistance to plant disease, better growth rate by modification

CC of the cell cycle pathway with plant growth regulators, increased rate of

CC homologous recombination, modified seed oil or protein yield and/or

CC content, improved yield by modification of carbohydrate, nitrogen or

CC phosphorus use and/or uptake, by modification of photosynthesis or by

CC providing improved plant growth and development under at least one stress

CC condition, improved lignin production or improved galactomannan

CC production. This sequence represents a bacterial polypeptide used in the

CC scope of the invention. Note: The sequence data for this patent did not

CC form part of the printed specification but was obtained in electronic

CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX

SQ Sequence 933 AA;

Query Match 39.8%; Score 1833; DB 8; Length 933;

Best Local Similarity 40.9%; Pred. No. 1.3e-143;

Matches 393; Conservative 178; Mismatches 281; Indels 110; Gaps 14;

QY 2 ENNDALDIGAVSSPPSQSEGVSTPLQVTFSPSFDNAASPAG----- 44

Db 11 EDNNS-----SSPVFPNPDSPV---PQLSSPALLFYSSSSQSDIYGRNNSQNLSSQEGN 61

QY 45 -RRAVROTP---TSAVRRRGRETDSARRRRRSRLGNSVYSSPVYDAGTGGTGGPVATP 100

Db 62 IRAAIGSPLNFPSSQRQNSDFVQSQGRQRISSASASGRSRVHSDLRDALPTSS- 120

QY 101 VYATPVGTGTPMGTSPFHRGTPOYKORSELGSGCKPLHRRRRSQSREPGRHSRSPSADGR 160

Db 121 -----SSLRGNQNRVHMRNDIHTSDLSSPRRIVDFDTR 155

QY 161 -----PSSAEPDPLTGGEYAVVGTNNVNPDLRAIRPLNHYR----- 200

Db 156 SGVNTLDTSSSSAPPSEASEP-----LRIWGTVNSIQECTTNFRNFLMSFKYFKRK 208

QY 201 -----SSAHDLSKYIQIIEETVREEDTLNIDMSDI--YDHDPLYAKIVYPLDI 250

Db 209 LDREEFINNTDBELYIKOLNEMRELGTNLNLDARNLLAYKQTELDYQLLNPOEV 268

QY 251 IPLLDTECEQVATSL-----LPTFE-KHIEARPENLKASVHMRLEINPSDIDKLVS 300

Db 269 ISINDQTIKDCMVSILVDNNDYDLDEIETFKYKVPYNVGVSCGKMRLEINPDIDKLINL 328

QY 301 KGMVIRCSSIPEIKGAPFKCLVCGHSPPLVTVKGRVEEBTRCEKPCAARNAMSLIHN 360

Db 329 KGLVLRSTPVPDMKVAFFKCNVCDHT--MAVEIDRGVIOEPARCEBIDCNPNMSLIHN 387

QY 361 RCTFANKQIVLQETPDPAIPEGETPHTVSMCLYNTMVDVAVKPGORIEVTGVFKAMAVVG 420

Db 388 RCSFADKQVVKLQETPDPFVDPGQTPHSISLCVDELVDSCRAGDRIEVTGTFRPIIRAN 447

QY 421 PNQRTLALYKTYIDCVHVKKSDRGRLQTEPDMEMDKENDMYAGVHESDTSSEANEAQ 480

Db 448 SRQVLSKLYTYVDVHVKKVSDKLDVDSTTIEQELMQNKVDHNEVEVRQITDQDLA 507

QY 481 XLKLSKLPGYIDRLSRSLAPSWELEDIKGLLQCLFGGKAKKIPSGASFRGDNVLLV 540

Db 508 KIREVAAREDLVSLARSIAPSIYELEDVKKGILLQLFGGNTKFTKGRYRGDINILLC 567

QY 541 GDPGTSKQLQYVHKIAPRGITYTSGRSSAVGLTAYVTKDPTETVLESGALVLSDRG 600

Db 568 GDPSTSKQLQYVHKITPRGVYTSRGSSAVGLTAYITRDVDTKQLVLESGALVLSDRG 627

QY 601 ICCIDFDFKMDNARSMLHEVMEQOTSVAKGGIIASLNARTSVLACANPSGRNARLS 660

Db 628 VCCIDFDFKMSDSVRSVLHEVMEQOTISAKAGIITLNARSSILASANPISGRNPNLP 687

QY 661 VIDNIQPLPTLLSFPDLIYMLDKPEQNDRLARHLVALHYEN--YEVSKQDADLQTL 718

Db 688 VTENIDLPPLLSRFDLVVLVDKVDKNDRELAKHLTNLYLEDKPEHISQDDVLPVBF 747

QY 719 TAYITYARQVHPITLSDEAAEDLINGVYEMKGNFPQSSKKVITATPROLESMTIRSEA 778

Db 748 TMYISAKEHTHPITTEAAKTELVRAYVGMKMGDDSDRSDEKRITATTTRQLESMTIRLAA 807

QY 779 LARFRFSVEKVDAAAEAVRLLDVALQOSATDHATGTTDMDLITTVGSASERIRANLIA 838

Db 808 HAKMKLVNVELEDVQEAURLIRSAIKDYATDPKTGKIDMNLVQTGKSVIQKQLQEDLSR 867

QY 839 ALRELIAADKIPSGSSSGLTKTSOLLEDIRSQSSVDVSLQDINKNALGSLQEGFLTVGH 898

Db 868 BIMVNLKQD-----ASDSMSFNLKQINEHSQDRVSSSDIQEALSRLQOQEDKVIVLG 923

XX WPI: 2004-786403/78.
DR N-PSDB; ADU05807.
XX New nucleic acid, and derived proteins, useful for diagnosis of bronchial
PT cancer and in screening for therapeutic and diagnostic agents.
XX
PS Claim 2; SEQ ID NO 518; 1381pp; German.
XX
CC This invention relates to a novel isolated nucleic acid associated with
CC bronchial cancer comprising 489 defined sequences given in the
CC specification. The invention may be useful for the production of
CC compounds with a cytostatic activity through the inhibition of expression
CC or activity of tumour-associated proteins. The novel DNA sequences and
CC the proteins/peptides encoded by them are used for detecting bronchial
CC cancer or determining the risk of developing it and to screen for
CC specific binding partners of the DNA or protein sequences, where the
CC binding partners are potentially useful as agents for treating or
CC diagnosing bronchial cancer. The DNA or protein sequences can also be
CC used for prognosis, detection of metastases and for secondary treatment
CC (of tumours that have been stabilised or are no longer detectable).
CC Detecting abnormal expression of the DNA sequences provides early
CC diagnosis of bronchial cancers. The present sequence is that of a protein
CC encoded by a novel bronchial cancer-associated human gene sequence of the
CC invention.
XX
SQ Sequence 924 AA;
Query Match 39.4%; Score 1817.5; DB 8; Length 924;
Best Local Similarity 43.7%; Pred. No. 2.5e-142;
Matches 404; Conservative 159; Mismatches 275; Indels 86; Gaps 19;
QY 18 SOSEGVSTPLP-QVTSFSPDAAAPVA-----GRRVAVROTPTSASVRRR-GRETDSARRR 70
DB 42 SREAGPARACPRAGTGS--TWSSPASTSPRSGRGRATPAQTFRSEADARSPSORRG 99
QY 71 SRSRSLG-----NSVSSPYDAGTGTGP-GTPVATPV-YATPVGPT 109
DB 100 EDSTSTGELQPMPTSPGVDLQSTAAQDLVFSPPQMHSSAIPLDVDSPLTYGTSSRV 159
QY 110 MGTPT-SFHRTGTOYKORSGLGQKPLHRRRSQSRQREPHSPRSPADGRPSSEAE 168
DB 160 EGTPTSGVGRTPT-VQRPDLGSAQKGLQVDLOS-----DGAADIVAS 202
QY 169 DTLGGYAYVWCTNNIPDVLRIRREPLHNY-----RSSAHLNSK-YIQIIEVTE 219
DB 203 EQLGQKLVINGDTGVNVAACKENFQRFIDFLAKEBENVGIDITELPWLQRLGEINV 262
QY 220 REEDTLNIDMSDIYDHPDLYAKIVRYPLDIIPLLDTECEVATSLLP--TFEKHTEARP 277
DB 263 IGEQFLVNCERHKSFDKNLYRQLISYQPEVPTFDMAVNEIPFDYRPSILHEQIOVRP 322
QY 278 FNLKASVHMRNLNPSDIDKLVSVMGVRICSSIIPEIKGAFKCLVCGHSPPLVTVVGR 337
DB 323 FNALKTKMNLNPNEDIDQITISGMVIRTSQLIPBQWQAFQOCVCAHT-TRVEMDRGR 381
QY 338 VEEPTCEKPECAARNAMSLIHRNCTFANKQIVRLQETPDPAIPEGTPTVSNCLNTMV 397
DB 382 IAEPSVCGR--CHTTSMALIHNRSLFSDQMILQESPEMDMPAGTPTHTVILFAHNDLV 439
QY 398 DAVKPGDRTEVTVGFKAMAVRVPQNRKTRALYKTYIDCVHVKKSRGRLOTEDEPMEMDK 457
DB 440 DKVQPGDRVNVGTIYRAVPVPRVNVSVKSVKTHIDVHYRKTAKRLH----- 490
QY 458 ENDMYAGYHSDTSEANEAKIQKJELSKLPGIYDRLSRLSPSIWELEDIKKGLLQCL 517
DB 491 -----GLDEEARQKLFSEKRVKLLKSLRKPDIYERLASALAPSIYEHEDIKKGILLQL 544
QY 518 FGGKAKKI--PSGASFRGDINVLVGPDTGSKQLQYVHKIAPRGIYTSGRSSAVGLT 575
DB 545 FGGTRKDFSHTRGKFRABINILLCGDPTGTSKQLLQYVNVPRGQYTSKGSSAVGLT 604
QY 576 AYVTKDPETRETVEESGALVLSRGICCIDDEFKMSDNARSMLHEVMEQQTVSVAKGII 635

DB 605 AYVWKDPETROLVLTQGTALVLSDNIGICCIDBFDKMNSTRSVLHEVMEQQTLSIAKAGII 664
QY 636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYMLDKPDRQNDRLAR 695
DB 665 COLNARTSVLAAANPESQWPKTTIENIQLPHTLLSRFDLIYMLDKPDRQNDRLAR 724
QY 696 HLVALHYENYVSKODALDLOTLYAYITVAROHVHTLSDEAAEDLINGVEMOKGNPP 755
DB 725 HLVALYQSEQAEBELDMAVLKDYIAYAHSTIMFRLSEASQALIEAYVDMRK-----I 780
QY 756 GSSKKVITATPROLESMTIRISEALARMFSEVVEKVDAEAVRLLDVALQQSATDHATCT 815
DB 781 GSSRGWASVPRQLESIRLAEAHAKVRLSNKVEAIDVEAKRLHREALKQSATDPTGTI 840
QY 816 IMDDLITTVGSASERRIRANLIALRELITADKISPGSSSGLTKTSQLEDIRQSSVDVSL 875
DB 841 VDISILTYGMSATSRKKEELAEALKKLI---LSKGTPLAKYQQLFEDIRQSSDIATK 897
QY 876 QDINKNALGSLQGEGLTVHGDIVK 899
DB 898 DMFEERLALADDDFLVTGKTVR 921
RESULT 12
ADR69821
ID ADR69821 standard; protein; 863 AA.
XX AC ADR69821;
XX DT 02-DEC-2004 (first entry)
XX DE Human minichromosome maintenance-4 (Mcm4) protein.
XX KW synthesis-period integration effect; cell cycle; phosphorylation;
KW minichromosome maintenance-4; Mcm4; Chk-1; cancer; Cdk-2; cytostatic;
XX OS Homo sapiens.
XX PN JP2004248594-A.
XX PD 09-SEP-2004.
XX PF 20-FEB-2003; 2003JP-00043174.
XX PR 20-FEB-2003; 2003JP-00043174.
XX PA (MITU) MITSUBISHI CHEM CORP.
XX DR WPI; 2004-656246/64.
XX PT Screening synthesis-period integration effect releasing substance of
PT cell, useful in treating cancer, by culturing cell with test substance
PT and selecting substance suppressing phosphorylation of minichromosome
PT maintenance-4 protein.
XX PS Example 1; SEQ ID NO 1; 30pp; Japanese.
XX CC The invention relates to a novel method for screening a substance capable
CC of releasing the synthesis-period integration effect of a cell. The
CC method involves culturing the cell in the presence of the test substance,
CC where the cell cycle has been stopped by the synthesis period, and
CC selecting the substance, which has the capability to suppress
CC phosphorylation of minichromosome maintenance-4 (Mcm4) protein, without
CC inhibiting the function of the Chk-1 protein in the cell. The invention
CC further comprises: a substance obtained by the method; a therapeutic drug
CC for treating cancer, comprising the substance obtained by the method, and
CC a DNA reproduction inhibitor; a medical agent for increasing the
CC therapeutic effect of treating the cancer by DNA reproduction inhibitor,
CC comprising the substance obtained by the method as an active ingredient;
CC and a medical agent for releasing the synthesis period integration effect
CC of a cell, comprising the inhibitor of the Cdk-2 protein as an active

[illegible]

RESULT 14
ABM80987
ID ABM8
XX
AC ABM8

18-NOV-2004	(first entry)
Tumour-associated antigenic target (TAT) polypeptide PRO81673, SEQ.2545.	
Tumour-associated antigenic target; TAT; human; overexpression; cancer;	
tumour; diagnosis; cell proliferative disorder; breast cancer;	
colorectal cancer; lung cancer; ovarian cancer; liver cancer;	
central nervous system cancer; bladder cancer; pancreatic cancer;	
cervical cancer; melanoma; leukaemia; hybridisation probe;	
chromosome identification; chromosome mapping; gene mapping;	
gene therapy; cytostatic.	
Homo sapiens.	
WO2004030615-A2.	
15-APR-2004.	
29-SEP-2003; 2003WO-US028547.	
02-OCT-2002; 2002US-0414971P.	
(GETH) GENENTECH INC.	
Wu TD, Zhang Z, Zhou Y;	
WPI; 2004-347921/32.	
N-PSDB; ACN38801.	
New tumor-associated antigenic target polypeptides and nucleic acids,	
useful in preparing a medicament for treating or detecting a	
proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
prostate cancer or tumor.	
Claim 12; SEQ ID NO 2545; 7273pp; English.	
The invention relates to human tumour-associated antigenic target (TAT)	
polypeptides, and their related nucleic acids. The TAT polypeptides are	
overexpressed in cancer tissues compared to normal tissues, and may thus	
serve as effective targets for the diagnosis and treatment of cancer in	
mammals. The invention also relates to nucleic acid and polypeptide	
sequences at least 80% identical to the TAT nucleic acids and	
polypeptides; expression vectors and host cells comprising a TAT nucleic	
acid; an antibody specific for a TAT polypeptide; a peptide or organic	
molecule which binds to a TAT polypeptide; fusion proteins comprising a	
TAT polypeptide; and methods and compositions for the treatment or	
diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
antibodies, antagonists, binding molecules and compositions are useful	
for diagnosing or treating a cell proliferative disorder associated with	
increased TAT expression, particularly cancers such as breast cancer,	
colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder	
cancer, pancreatic cancer, cervical cancer, cancers of the central	
nervous system, melanoma and leukaemia. TAT nucleic acids may further be	
used as hybridisation probes, in chromosome and gene mapping, in	
chromosome identification and in gene therapy. The present sequence	
represents a TAT polypeptide of the invention	
Sequence 863 AA;	

```

Query Match      39.3%; Score 1810.5; DB 8; Length 863;
Best Local Similarity 44.1%; Pred. No. 8.6e-142;
Matches 400; Conservative 158; Mismatches 266; Indels 83; Gaps 18

QY 30 VTSPSFDNAASPV-AGRAVQTPTPSVRRR-GRETDSARRRSRSRSLG----- 77
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 1  MSSP-----ASTPSRRGSRGRATPAQTFRSDARSPSQRRRGEEDSTSGELQPMPTSPG 56
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 78 -----NSVYSPDYDAGTGPPT-GTPTATPV-YATPTVGTGPMGTP-SFHRGTGTPQYKOR 125
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 57 VDLQSPAQAQDVLFSPPQPMHSSAIELDFDVSSPLTYGTPTSSRVEGTPTPSGVGTPT-VRQR 115
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 126 SELGSQGKPLHRRRRSQSPGCHRSPPSREPSADGRPPSSAEPDDTLGGBYAYVTGWNVNI 185

```

```
Db 116 PDLSAGKGLQVDLQS-----DGAADIVASEQSLQKLVINGTDVNV 159
Qy 186 PDVLAIRFRFLNY-----RSSAHLNKG-YIQIIEETVEREDTLNIDMSDIYDHD 236
Db 160 AACKENFORFLQRFIDPLAKEENGVGIDITEPLYNQRLGEINVIGEPFLNVCEHIKSF 219
Qy 237 PDVAKIVRYPLDIIPLDTECOEVATSLP--TFEKHIEAPFNKASVHMRELNPDSI 294
Db 220 KNLVQLSYPQEVITPTDMVNEIPFRYPDSILHQLQVRFPNALATKNMNLNPEDI 279
Qy 295 DKLVSVMGIVRCSSIIPEIKGAFFKCLVCGHSPPLVTWVKGRVEPTRCEKPECAARNA 354
Db 280 DOLITISGWIRTSQILPEMQEAFQCVQCAHT-TRVEMDRGRIAPESVCGR--CHTHHS 336
Qy 355 MSLIHNRTCTPANKQIVRLQETPDALPEGETPHTVSMCLYNTWVDAVKPGDRIEIVGVKPA 414
Db 337 MALIHNRLSFDKQMKIQESPEMPAGCTPHTVILFAHNDLVKVPQGDVNVVTGIYRA 396
Qy 415 MAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQETDPEMDKENDMYAGHESDTSEA 474
Db 397 VPVRNVPRVSNVSKYKTHIDVIHYRKTDAKRLH-----GLDEEABQKLF 441
Qy 475 NEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEIKGLLQQLFGGKAKKI--PSGASFR 532
Db 442 SEKRVELLKELSRKPDIIYERLASALAPSIYEHEDIKGIILLQFGGTRKDSHTGRGKFR 501
Qy 533 GDINVLVGDPTGSKQLQVYVHKIAPRIYVTSGRGSSAVGLTAYVTKDPETRETVLBSG 592
Db 502 ABEINLLCGDPTGSKQLQVYVLPVPRQVTSGRGSSAVGLTAYVTKDPETRQLVLQTG 561
Qy 593 ALVLSDRGICCIDPEPKKSDNARSMLHEVMEQOTVSVAKGGLIASLNARTSVLACANPSG 652
Db 562 ALVLSDRGICCIDPEPKKNESTRSVLHEVMEQOTLSIAKAGIICQLNARTSVLAANPIE 621
Qy 653 SRYNARLSVIDNIQPLPTLLSRFDLIYMLDKPQDNDRLARHLVALHYENYEVSKODA 712
Db 622 SQWNPCKTIIENIQPLHLLSRFDLIIFLLDPQDAYDRRLAHLVALHYEQSEQAEEEL 681
Qy 713 LDLOTLTAYITVARQHVHTPSDEAABLNGVEMRQKGNPFGSKKIVTATPRLQESM 772
Db 682 LDMVLKDVIYAHGHTIMPRLSEASQALIEAYVDMRK----IGSSRGWVSAYPRLQESL 737
Qy 773 IRISALARMPESEVVEKVDAAEAVRLDLVALQOSATDHATCTIDMTLITGVSAERIR 832
Db 738 IRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTGTMSATSRK 797
Qy 833 RANLLAALRELITADKISPGSSGLTKTSOLLEDIRSQSSVDVSLQDIKNALGSLQCEGLT 892
Db 798 KEELAEALKKLI---LSKGTPLALKYQQLFEDIRQSDIATYKDMFEALRALADDDFLT 854
Qy 893 VHGDIVK 899
Db 855 VTGKTVR 861

RESULT 15
ID ABP73643
XX ABP73643 standard; protein; 910 AA.
XX AC ABP73643;
XX 30-JAN-2003 (first entry)
DT Candida albicans essential protein SEQ ID NO 7480.
DE Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX Candida albicans.
XX OS WO200253728-A2.
XX PN
```

```
XX 11-JUL-2002.
PD 26-DEC-2001; 2001WO-US049486.
XX 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX (ELIT-) ELITRA PHARM INC.
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
PI WPT; 2002-566694/60.
XX N-PSDB; ABZ33193.
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX Claim 44; SEQ ID NO 7480; 167pp + Sequence Listing; English.
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungal
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX Sequence 910 AA;
SQ
```

Query Match 38.1%; Score 1754; DB 5; Length 910;
Best Local Similarity 42.0%; Pred. No. 5,1e-137;
Matches 397; Conservative 155; Mismatches 259; Indels 134; Gaps 20;

```
Qy 3 NNDALDIGAVSPYPSQSE---GVSTPLPQVTPSPFDNAASPVAGRAVRQPTSAV--- 56
Db 45 SNPGSDIGNFSQSQSQRNDISSPLHYTSSAQ-----PTSDIGGF 87
Qy 57 --RRRGRETDSARRRRRSRSLGNSVYSSPYDAGTPTGTPVATPVATPVGTPMGTPS 114
Db 88 DSQRSARVQDVGRIMERAQRSDVTVDSVSP----- 117
Qy 115 FHRGTPQYKQSELGSGQKPLHRRRRSQSRPGRHSRSPRESADGRPSAEPDPTLGE 174
Db 118 -----QRS-----RRYFTQGRNGPNSLNSSTSAQ-FSTDPAEPNDE---P 154
Qy 175 YAYVGTWNVNIPDVLRAIRREL---HNYRSSA-----HDLSNKYIQIIEETVERE 221
Db 155 VRVWGTNVSQECESNIFRDLFLSPKYKIRELEEQALEPEDEHEL--YVNOJNNIIEUG 212
Qy 222 EDTLNIDMSDI--YDHPDPLAKIVRYPLDIIPLDTECOEV-----ATSL 265
Db 213 LTNLNLDAKLLSYPTSKLYVQLINYQEIIPIMDHTIKDCLQIINDANATTSPAQSK 272
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:35:21 ; Search time 47 Seconds
(without alignments)
1584.910 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALDIGAVSSPYPSQS.....LGSLLQGGFLTVHGDIVKRV 901

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/prodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/prodata/1/1aa/H COMB.pep.*
4: /cgn2_6/prodata/1/1aa/PCTUS COMB.pep.*
5: /cgn2_6/prodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4609	100.0	901	2 US-09-828-062-8	Sequence 8, Appli
2	965.5	20.9	879	2 US-09-949-016-8492	Sequence 8492, Ap
3	963.5	20.9	720	2 US-09-604-605-2	Sequence 2, Appli
4	963.5	20.9	720	2 US-10-101-080-2	Sequence 2, Appli
5	946.5	20.5	904	2 US-09-964-899-49	Sequence 49, Appli
6	939	20.4	753	2 US-09-949-016-10892	Sequence 10892, A
7	936	20.3	909	2 US-09-976-594-324	Sequence 324, App
8	932	20.2	1017	2 US-09-538-092-319	Sequence 319, App
9	919.5	20.0	895	2 US-09-538-092-1158	Sequence 1158, Ap
10	916.5	19.9	823	2 US-09-949-016-8339	Sequence 8339, Ap
11	817.5	17.7	544	2 US-09-248-796A-18637	Sequence 18637, A
12	734.5	15.9	495	2 US-09-248-796A-20007	Sequence 20007, A
13	662	14.4	331	2 US-09-248-796A-18673	Sequence 18673, A
14	474	10.3	435	2 US-09-248-796A-18814	Sequence 18814, A
15	391.5	8.5	173	2 US-09-248-796A-20019	Sequence 20019, A
16	357	7.7	429	2 US-09-248-796A-19091	Sequence 19091, A
17	246.5	5.3	163	2 US-09-248-796A-20018	Sequence 20018, A
18	216	4.7	248	2 US-09-248-796A-19090	Sequence 19090, A
19	199	4.3	353	2 US-09-248-796A-20008	Sequence 20008, A
20	152.5	3.3	1781	2 US-09-961-403-13	Sequence 13, Appli
21	151	3.3	265	2 US-09-248-796A-18812	Sequence 18812, A
22	150	3.3	1493	2 US-09-423-890-8	Sequence 8, Appli
23	149.5	3.2	1492	2 US-09-697-898-5	Sequence 5, Appli
24	149	3.2	1780	1 US-08-769-309A-5	Sequence 5, Appli
25	149	3.2	1780	2 US-08-994-570-5	Sequence 5, Appli
26	142	3.1	105	2 US-09-513-999C-7279	Sequence 7279, Ap
27	142	3.1	1492	2 US-09-697-898-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-828-062-8
; Sequence 8, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/09/828,062
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-09-828-062-8

Query Match	100.0%;	Score	4609;	DB	2;	Length	901;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	901;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MENNDALDIGAVSSPYPSQSEGVSTPLQVTSFDFNAASPVAGRAVRCQTPTSARRRG	60				
Db	1	MENNDALDIGAVSSPYPSQSEGVSTPLQVTSFDFNAASPVAGRAVRCQTPTSARRRG	60				
Qy	61	RETDSARRRRSRSLGNSVYSSPYDAGTGGTGTGPTVATPVATGTPMGTPSFHRTGTP	120				
Db	61	RETDSARRRRSRSLGNSVYSSPYDAGTGGTGTGPTVATPVATGTPMGTPSFHRTGTP	120				
Qy	121	QYKQSELGSGKPLHRRRRSQRSPGHRSPSREPSADGRPSASPDPTLGGEYAYVWG	180				
Db	121	QYKQSELGSGKPLHRRRRSQRSPGHRSPSREPSADGRPSASPDPTLGGEYAYVWG	180				
Qy	181	TNNVDPVLRATRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHPDLY	240				
Db	181	TNNVDPVLRATRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHPDLY	240				
Qy	241	AKIVRYPLDIIPLLDTECOVATSLTPTKEKTEARFPNLSKASVHMRNLNPSDIDKLVS	300				
Db	241	AKIVRYPLDIIPLLDTECOVATSLTPTKEKTEARFPNLSKASVHMRNLNPSDIDKLVS	300				
Qy	301	KGMVIRCSIIPEIKGAPFKCLVCGHSPLVTYVVKGRVBEPTCEKPECAARNAMSLIHN	360				

Db 301 KGMVIRCSIIPIEIKGAFFKCLVCHSPPLVTVVKGRVSEPTCEKPECAARNASLIHN 360
Qy 361 RCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTMVDVAVPGDRIEIVTGVFKAMAVRG 420
Db 361 RCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTMVDVAVPGDRIEIVTGVFKAMAVRG 420
Qy 421 PNQRTLALYKTYIDCVHVKSDRCRLQETDPMEMDKENDMYAGVHESDTSSEANEAKIQ 480
Db 421 PNQRTLALYKTYIDCVHVKSDRCRLQETDPMEMDKENDMYAGVHESDTSSEANEAKIQ 480
Qy 481 KLKELSKLPGIYDRLSRSLAPSIMELEDIKKGLLQCLFGKAKKIPSGASFRGDNVLLV 540
Db 481 KLKELSKLPGIYDRLSRSLAPSIMELEDIKKGLLQCLFGKAKKIPSGASFRGDNVLLV 540
Qy 541 GDPGTSKQLQYVHKIAPRGIIYTSGRSSAVGLTAYVTKDPETRETIVLESALVSDRG 600
Db 541 GDPGTSKQLQYVHKIAPRGIIYTSGRSSAVGLTAYVTKDPETRETIVLESALVSDRG 600
Qy 601 ICCIDFDMKSNABSMLEHVEWQOTVSAGKGIITASLNARTSVLACANPGSGRYNARLS 660
Db 601 ICCIDFDMKSNABSMLEHVEWQOTVSAGKGIITASLNARTSVLACANPGSGRYNARLS 660
Qy 661 VIDNIQLPPTLSRFDLIYMLDKPEQNDRLARHLVALHYENYEVSKQDALDQTLTA 720
Db 661 VIDNIQLPPTLSRFDLIYMLDKPEQNDRLARHLVALHYENYEVSKQDALDQTLTA 720
Qy 721 YITYARQHVHPTLSDEAEDLINGVYEMRQKNFPGSSKKVITATPROLESIRISEALA 780
Db 721 YITYARQHVHPTLSDEAEDLINGVYEMRQKNFPGSSKKVITATPROLESIRISEALA 780
Qy 781 RMRFSEVVEKVDAAEAVRLDVALQOSATDHATGIDMDLITTVGSASERIRANLLAAL 840
Db 781 RMRFSEVVEKVDAAEAVRLDVALQOSATDHATGIDMDLITTVGSASERIRANLLAAL 840
Qy 841 RELIADKISPGSSGLKTSQLEDDIRSSQSSVDVSLQDINKNALGSLQSGFLTVHGDIVKR 900
Db 841 RELIADKISPGSSGLKTSQLEDDIRSSQSSVDVSLQDINKNALGSLQSGFLTVHGDIVKR 900
Qy 901 V 901
Db 901 V 901

RESULT 2
US-09-949-016-8492
; Sequence 8492, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8492
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8492
Query Match 20.9%; Score 965.5; DB 2; Length 879;
Best Local Similarity 30.9%; Pred. No. 5.1e-76;
Matches 264; Conservative 143; Mismatches 249; Indels 197; Gaps 30;

Qy 130 SOGKPLHRRRSQSGREHRSRPSGADGRPSABPDDTLGGEYAVVWGTNNVNP-V 188
Db 38 SVGRPQSTARRF-SASPGAR-PLSPPSAAARPSAAMALKD-----YALEKGTGLRAREGV 90
Qy 189 LRA-----IRRFPLHNYRSASHDLNSKY 210
Db 91 LRGEPPGSHLLFIGRVLSENGYLRFVLKFVSYGFTFHVTEKVKFLQEFYQD-DELGKKQ 149
Qy 211 IQIITEV---EREDTINIDMSDIYHDDPL-----YAKIVRYPL-DIIP----- 252
Db 150 FKYGNQLVRLAHREQVALYVDLDDVAEDDPELVSDICENARRYAKLFADAVQVELLPQYKE 209
Qy 253 -----LLDTECO-----EVATSLLPTEKHEARPENLKASVH 285
Db 210 REVNVKVDLDVYIEHRLMMEORSRDPQVVRFPQOYPAELMRRRFLYFQG-PSSNKPV- 267
Qy 286 MRELNPSSDIDKLAVSVKGMVIRCSIIPEIKGAFFKCLVCG---HSPPLVTVVVKGRVSEPT 342
Db 268 IREVRADSVGKLVTVRGIVTRVSEVKPMVATYTCDCQGAETYP-----IQSPT 318
Qy 343 -----RCEKPECAARNA---MSLIHNRCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYN 394
Db 319 FMPLIMCPQSQRCQNRSGRLYLQTRGSRFIKFOEMKMQEHSQDQVPVGNIPRSITVLVEG 378
Qy 395 TMVDVAVPGDRIEIVTGVFKAMAVRG-----PNO-RT-----L 426
Db 379 ENTRIAQPGDHVSVTGIFLPI-LATGFRQVVRKEQNPQNRTQGGFVKLLGNPVMGKL 437
Qy 427 RALY-----KTYIDCVHVKSDRCRLQETDPMEMDKENDMYAGVHESDTSSEA 473
Db 438 KGLHLYPGIFLPKGLLSETYLEAHRIVK-----MNKSEDDSEG----- 475
Qy 474 ANEAKIOKLKELSKLPGIYDRLSRSLAPSIMELEDIKKGLLQCLFGKAKKIPSGASPRG 533
Db 476 AGELTRBELRQIAE-EDFYEKLAASIAPIEYGHEDVKKALLLLVGG-VDQSPRGMKIRG 533
Qy 534 DINVLVGDPTGTSQQLQYVHKIAPRGIIYTSGRSSAVGLTAYVTKDPETRETIVLESQA 593
Db 534 NINICLMDGPVAKSQLLSYIDRLAPRSQYTTGRGSSGVGLTAAVLRSVSGELTLEGA 593
Qy 594 LVLSDRGICCIDFDMKSNABSMLEHVEWQOTVSAGKGIITASLNARTSVLACANPGSGS 653
Db 594 LVLDQGVCCIDFDMKABADRTAIHEVMEQOTISIAKAGILITTLNARCSILAAANPAYG 653
Qy 654 RYNARLSVIDNIQLPPTLSRFDLIYMLDKPEQNDRLARHLVALHYENYEVSKQ-DA 712
Db 654 RYNPRRSLEQNIQLPAALLSRFDLLWLITQDRPRDNDRLAQHITYVHQHSRQPPSPQFEP 713
Qy 713 LDLOTLTAYITYARQHVHPTLSDEAEDLINGVYEMRQKNFPGSSKKVITATPROLES 772
Db 714 LDMKLMRYIAMCREK-QPMVFPESLADYITAYVEMRSEA---WASKDATYTSARTLLAI 769
Qy 773 IRISSEALARMFSEVVEKVDAAEAVRLDVALQOSATDHATGIDMDLITTVGSASERIR 832
Db 770 LRLSTALARLWMVDVEKEDVNEAIRLWMSKDSILGD-----KGQTARTQRP 817
Qy 833 RANLLAALRELIA 845
Db 818 ADVIFATVRELVS 830
RESULT 3
US-09-604-605-2
; Sequence 2, Application US/09604605
; Patent No. 6421613
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/09/604,605
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/143,222


```

; Sequence 49, Application US/09964899
; Patent No. 6900367
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-964-899-49

Query Match      20.5%; Score 946.5; DB 2; Length 904;
Best Local Similarity 29.0%; Pred. No. 2.6e-74;
Matches 262; Conservative 157; Mismatches 344; Indels 139; Gaps 23;

Qy      29  QVTPSPFNAAPVAGRAVRCTPTSAVRRRGRETDSARRRRSRKSLGNSVSPYDAG 88
Db      2  QESSEFTWASP-AQRAGNDPLTSSPCRSRRDAL--TSSPGDLD-----PPFDE 52

Qy      89  TPGTGTPTVATPVATPVGTPM---GTPSFHRTGTPYKQSELG-----SQG 132
Db      53  SEGILGT--EGLEBEDGEELIGDMRDYRAIPELDAYEABGLADDEDVEELTASOR 110

Qy      133 KPLHRRRSQREPG-----HRSPPRESADQPSAESPD-----168
Db      111 EAAERAMQRDEAGRLGRMRGLLYDSDEDEERPAKRQVRATEDGEDEEMIES 170

Qy      169 ----DTLGEYAYV---GTNVNIPDVLRATRRFLHNY-RSSAHDLSNKYIQIIEETVE 219
Db      171 IENLEDLKGHSVREWVSNAGPRL---EIHRRFNFLRTHVDVSHGNV---FKERISDMCK 224

Qy      220 REEDTLNIDMSDIYDHPDLVAKIVRYPLDIPLDTECQEVATSLPTFEK---HIEAR 276
Db      225 ENRESLVNVEDLAAREHVLAYFLPEAPAEALLQIFDEAALEVVLWAMYPKYDRITNHHVR 284

Qy      277 PPNLKASVHMRELNPSDDKLVSVKGMVIRCSIIPEIKGAPFKCLVCHGHPPLVTVVKG 336
Db      285 ISHLPLVLEELSLRQLHLNQLRTSGVVTCTGVLPQLSWKYNKCNKCNFVLGPPCQSQN 344

Qy      337 RVEEPTRCBKPCAARNAMSLIHNRCFTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTM 396
Db      345 QEVKPGSC--PCQAGPFEVNMETIYQNYQIRIQESPGKVAAGRIPRSKDAILLADL 402

Qy      397 VDAPKPGRIEIVGVFKAMAVRVGNPQRTLRA--LYKTYIDCVHVKKSDR----GRLOTE 450
Db      403 VDSCKPGDEIELTGIYHN---NYDGLSNTANGFPVFAIVILANHVAKKONKVAANGELTDE 459

Qy      451 DPMEMDKENDMYAGVHESDTSAAEAKIQKLSKLPGLIYDRLSRSLAPSISWELEDIK 510
Db      460 D-----VOMITSLSKDQOQIGEKIFASIAFSPYIHGEDIK 492

Qy      511 KGLLQCLFGGKAKKIPSGASFRGDIINVLLVGPDPGTSKQLQYVHVHKAIPRGITVSGRSS 570
Db      493 KGLAALFPGEPNPGGKHVKRGDINVLLCGDPGFAKSKQLKIETKVSRAIFFTGQGAS 552

Qy      571 AVGLTAYYTKDPETRETLESALVLSRIGICIDEFDKMSDNARSMLHVMEEQQTQSVVA 630
Db      553 AVGLTAYYQVRHPVSREWTLKAGALVLAADRGVCLIDFEDKMDQDRTSIEHAMEQQSIS 612

Qy      631 KGGIITASLNARTSVLACANPSSGRNARLSVTDNLTQLPPTLLSRFDLIYMLDKPEDND 690
Db      613 KAGIVTSLQARCTVIAAANPIGGRYDPSLFTFSENVDLTPTIISRFDLICVVRDTPVDQD 672

```

```

Qy      691 RRLARHLVALHYENYEVSKQ-----DALDLQTLTAYITTAARQHVH 730
Db      673 EMLARFVGVSHVRHHPSPNKBEEGLANGSAABPAMPTVGVPELPQEVLLKVIYAKERVH 732

Qy      731 PTLSDAAEDLINGVEMRKGNPFGSKKVVITATPROLESIMIRISEALARMRFSEVEK 790
Db      733 PKNQMDDQKVAKMYSDLRKESMATGS----IPITVRHIESMIRMAEAAHRIHLRDYVIE 788

Qy      791 VDAAEAVRLDLVALQQSATQHTATCTIDMDLTTGVSASERIRAN---LLAALRELIADK 847
Db      789 DDVNNAIRV-----MLESFIDTQKFSVMRSMRK-T-FARYLSFRDNNEILLFILKOLVAEQ 843

Qy      848 IS 849
Db      844 VT 845

RESULT 6
US-09-949-016-10892
; Sequence 10892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10892
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10892

Query Match      20.4%; Score 939; DB 2; Length 753;
Best Local Similarity 30.9%; Pred. No. 8.7e-74;
Matches 232; Conservative 156; Mismatches 290; Indels 72; Gaps 18;

Qy      169 DTLGEYAYVGTNVNIPDVLRATRRFLHNYR--SSAHDLSNKYIQIIEETVEREEDTLN 226
Db      32 DSFGGD-AQADGQARKSQLQRRFKFLRQYRVGTDRTGFTFKYRDELKRYNLGEYWIE 90

Qy      227 IDMSDIYDHPDLVAKIVRYPLDIPLDTECQEVA---TSLPTFEKHIEARPNLKAS 283
Db      91 VEMEDLASFDLEDLADLYLKQPAEHLQLLEAAKEVADEVTRPSPGSEVLODIQVMLKSD 150

Qy      284 V---HMEELNPSDIDKLVSVKGMVIRCSIIPEIKGAPFKCLVCHGHPPLVTVVKG--RV 338
Db      151 ASPSISRLSKDMMSHLVKIPGIIIAASAVRAKATRIQCSRCNTLTNTAMRRGLECY 210

Qy      339 BEPTRCE-----KPECAARNAMSLIHNRCFTFANKQIVRLQETPDPAIPEGETPHTVSMCLY 393
Db      211 ALPRKNTDQAGRPCKPL-DPYFIMPDKCKVDFQTLKLQELPDAVPHGEMPRHMQLYCD 269

Qy      394 NTWDAVKPGRIEIVGVFKAMAV-----RVGNPQRTLRYLYKTYIDCVHVKKSDR 444
Db      270 RYLCDDVVPVGNRVITMGIIYSIKKFLGTTSRGRDRVGVGIRS-----SYIRVLGIQ---- 319

Qy      445 GRLOTEDEPMEDKENDMYAGVHESDTSAAEAKIQKLSKLPGLIYDRLSRSLAPSIW 504
Db      320 -----VDTDGSGRSPAG-----AVSPQEEEFERLAALPNVETVSKSIAPSLP 363

Qy      505 ELEDIKKLLCOLFGGKAKKIPSGASFRGDIINVLLVGPDPGTSKQLQYVHVHKAIPRGITV 564
Db      364 GGTDMKKAIACLLFGGSRKRLPDGLTRRGDINLLMDPGTAKSQLKAFVEKSPIGVYT 423

```


Db 613 ARCTVIAANPIGGRYDPSLTFSENVDLTETPIISRPDILCVVRDTPVQVDEMLARFVVG 672
Qy LHYENYYSKO-----DALDQLTAYITVARQHVPTLSDEAAE 739
Db 673 SHVRHPSKEEGLANGSAEPAMPNTYGVPELPQEVKKYIYAKERVHPKLNQMOD 732
Qy 740 DLINGYEMROKGNPPGSKKVTATPQLRSMIRISEALARMSEVVEKVDAAEAVRL 799
Db 733 KVAKMYSDLRKESMATGS-----IPITVRHIESMIRMAEBAHARIHLRDYVIEDDVNMAIRV 788
Qy 800 LDVALQOQSHATCTIDMDLITTCVSASERIRAN---LLAALRELIADKIS 849
Db 789 ----MLESFIDTPQKFSVMRSRMTK-FARYLSFRDRNNELLFLKQLVAEQVT 836

RESULT 10
US-09-949-016-8339
; Sequence 8339, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8339
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8339

Query Match 19.9%; Score 916.5; DB 2; Length 823;
Best Local Similarity 32.3%; Pred. No. 1e-71;
Matches 252; Conservative 114; Mismatches 282; Indels 133; Gaps 22;

Qy 195 FLHYRSSHADLNSKYQIIEBTVEREEDTLNDSMDIYDHPDLVAKIVRVPLDILPL 254
Db 35 FLEEFQSS--DGEIKYQLAEELIRPERNLVVSFVLDLQFNQQLSTTIQEEFYRVYPL 92
Qy 255 DTEQEVATSLPTPEKHIEARPF-----NLKASVHMRLENPSPDIDKLVSQGMVI 305
Db 93 ---CR-----ALKTFVKORKEIPLAKDFVAFQDLPTRHKIRELTSSRIGLLTRISQGV 144
Qy 306 RCSSIIPEIKGAFFKCLVCGHSPPLVTVVKG-----RVBEPTRCEKPECAARNAMSLIHN 360
Db 145 RTHVPHELVSQTFCLDC-----QTVIRDEVOQFQKTYQPNICRNPVCANRRFLDTN 198
Qy 361 RCTEANKQIVRLQETPDATPEGSTPHTVSMCLYNTMVDVAVKGDRIEVTGV----- 411
Db 199 KSRFVDFQKRVIOETQAEIPLRGSIPRSLRVLRAEAESAQAGDKCDFGTGLIIVPDVSK 258
Qy 412 FRAMAVRVGNORT-----LRAL-----YK-TYIDCVHVKKSDRGRLQTED 451
Db 259 LSTPGARAEVNSRVSGVDGYETEGIRGLRGLVRLDLSYRLVFLACC----- 304
Qy 452 PMEMDKENDMYAGYHSDTSEANEAKIQ-----KLKELSKLPGIYDRLSRSLAPSTW 504
Db 305 ---VAPTNPRFGKELRDESEQTAESIKNQMTVKEMEKVPFMSQDNLYHNLCSTLFPPTIH 361
Qy 505 ELEDIKGLLQCLFGGKAKKIPSGASFRGDIIVLLVGDPTGSKSOLLQVYVHKIAPGIYT 564
Db 362 GNDEVKRGVLLMFGGVPKPTTGBGTSLRGDINVCIVGDPSTAKSQFLKHVEFSRPAVYT 421
Qy 565 SGRGSSAVGLITAVTVKDPETRETIVLESSGALVLSDRGICCIDBFDKMSDNARSMLHEVMEQ 624

Db 422 SGKASSAGLTAAVVRDEESHEFVIEAGALMLADNGVCCIDEFDRQVDRDQVVAIHEAMEQ 481
Qy 625 QTVSAVGIIITASNARTSVLACANPSSGRYNARLSVIDNIQLPPTLSRFDLIYMLDK 684
Db 482 QTSITRAGVKATLNARTSILAAANPISGHYDRSKSLKQNLNSAPINSRFDLPFILVDE 541
Qy 685 PDEQNDRLARHLVALHYENTEVSKQDALDQLTAYITVARQHVPTLSDEAAEDLING 744
Db 542 CNEVTDVAIARIVDLH-SRIEESIDRVYSLDDIRRYLLFARQ-FPKIKISESEDFIVEQ 599
Qy 745 YVEMROKGNFPSSKKVITATPROLESIRISEALARMSEVVEKVDAAEAVRL----- 800
Db 600 YKHLRQDQ-SGVTKSSWRITVROLESIRISEALARMSEVVEKVDAAEAVRL----- 658
Qy 801 ----DVALQQS-----ATDHATGTI-----DMDLITTVGSA----- 827
Db 659 IRVETPDVNLQDEBEIQMEVDEGAGGINGHADSPAPVNGINGYNEIDINQESAPKASRLRG 718
Qy 828 -SERIRANLLAA-LRELIADKISPGSSGLKTSQ-----LEDIRSQSSVDVSLQDIKVA 881
Db 719 FSEYCRISNLIVLHLRKVEESE-----DSALKRSELVNVWYLKEISEIDSEBELINKKRI 774
Qy 882 L 882
Db 775 I 775

RESULT 11
US-09-248-796A-18637
; Sequence 18637, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18637
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18637

Query Match 17.7%; Score 817.5; DB 2; Length 544;
Best Local Similarity 32.2%; Pred. No. 2.9e-63;
Matches 188; Conservative 108; Mismatches 181; Indels 107; Gaps 9;

Qy 308 SSIIPKIGAFKCLVCGHSPPLVTVVKGVRVEEPTRCPEKPECAARNAMSLIHNRCCTFANK 367
Db 1 TGVFPQLKVKVPCDLCKG--VVLVPYVDSNTEVKSIFCTNCSQSGKPFKLNSEKTLRYNY 58
Qy 368 QIVRLQETPDATPEGSTPHTVSMCLYNTMVDVAVKGDRIEVTGVFKAMAVRGNQRTLR 427
Db 59 QRITLQEPAGTVPAGRLPRHREIVLLSDLVADVAKPEDIEVTGIYK-----NNYDGNLN 112
Qy 428 A-----LYKTYIDCVHVKKSDRGRLQTEPDMEMDKENDMYAGYHSDTSEANEAKIQKL 482
Db 113 AKNGFPVATILEANSIRR-----KESSAFMG--GNNLVNMTTEBIRF 155
Qy 483 KELSKLPGIYDRLSRSLAPSIWELEDIKKGLLQCLFGGKAKKIPSGASFRGDIIVLLVGD 542
Db 156 RKLSEKGIIDIKIASMAPSIYGHDKITAIACSLFGGVKDVNGKLSIRGDIIVLLVGD 215
Qy 543 PGTSSQQLQYVHKIAPRGIYTSGRGSSAVGLITAYITKDPETRETIVLESSGALVLSDRGIC 602
Db 216 PGTAKSQILKYAEKTASRAVFATQGSAGVLTASVVRKDPITREMTLQGGALVLDKGTG 275


```
; Sequence 18814, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18814
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18814

Query Match      10.3%; Score 474; DB 2; Length 435;
Best Local Similarity 28.5%; Pred. No. 5.2e-33;
Matches 135; Conservative 84; Mismatches 170; Indels 84; Gaps 14;

QY 161 PSSEAPDPTLGGYAYVGTNNIP-----DVLRAIRRFLLHYRSSHADLNSKYIQII 214
Db 10 PPPTADQDDT-----NQPLDAIFGDRVRRFOEFLDRIDSNT---GIDYRSII 53

QY 215 ETVREEDTLNIDMSDIYDHPDLYAKIVRYPLDIPLDTECQEVATSL-----PTFE 270
Db 54 KDLMIKSPRLSVSDEIRDFREFWLGLLNQPADYLACERALRDYVLAIYDQDPSPF- 112

QY 271 KHEIARP-----FNLKASVHMRLEPNDSID-----KLVSQGMVIRCSIIPEI----- 314
Db 113 PHDSYDPNQOYYLSFGAFGGHSLFPRSIDSSYLSKMSVIEGIVTFASLVRPKVIRSVHY 172

QY 315 ----KGAPFKCLVCGHSPPLVTVVKGVRVEBPTRCEKPECAARNAMSLIHNRCCTFANKQIVR 371
Db 173 ASKTGRFYAREYRDQTTSPDAIATPAI-YPT-----EDMEGNKLITTEGYSTYRDYQKIS 226

QY 372 LQETPDALPEGETPHTVSMCLYNTMTDAVAKPGDRIEVTGVPKAMAVRCPNQRTLRALYK 431
Db 227 VQEMPETAPPGLPRSDVILDDDLVLTGPGDRVQIVGVYRAGLGAANNS-----SSFK 281

QY 432 TVIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSAAANE-----AKIQKLKELSK 487
Db 282 TVI-----LSNSVYL-LHARSTGVASQEKLTQDIDRNINKLAK 318

QY 488 LFGIYDRLSRSLAPSIWELEDIKKGLLQCLFGGKAKKIPSGASFRGIDINVLVGDPTGSK 547
Db 319 DRKIPDILSRSLAPSIYGFYDIKKAVLLMMWGMGVEKNLNGTHLRGDIINILWVGDPTAK 378

QY 548 SOLLQVHKIAPRGIVTSGRGSAVGLTAYVTKDPETRETIVLESGLALVLSDRG 600
Db 379 SQVLRPVLNTASIAIATTCRGSSGVGLTAAVTTDKETGERRLEAGAILTERRAG 431
```

RESULT 15

```
US-09-248-796A-20019
; Sequence 20019, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
```

```
; SEQ ID NO 20019
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20019

Query Match      8.5%; Score 391.5; DB 2; Length 173;
Best Local Similarity 47.0%; Pred. No. 2.2e-26;
Matches 70; Conservative 40; Mismatches 36; Indels 3; Gaps 3;

QY 262 ATSLLPTEKHI-EARPFNLK-ASVHMRLEPNDSIDKLVSQGMVIRCSIIPEIKGAPP 319
Db 17 AQSKLDEIETNVYTIIRPYNNVLVEKGIRELNPNDIDKLVSQGLTLRSTSIIPDMKVAFF 76

QY 320 KCLVCGHSPPLVTVVKGVRVEBPTRCEKPECAARNAMSLIHNRCCTFANKQIVRLQETPDAL 379
Db 77 RCNACGHTVG-VEIDRGVISEPTKCPREVCGQTNSMVLHNRRSSPSDKQVLIKQETPDIV 135

QY 380 PEGETPHTVSMCLYNTMTDAVAKPGDRIEV 408
Db 136 PDGOTPHSINLCVYDELVDSCRAGRVDL 164

Search completed: December 7, 2005, 17:44:36
Job time : 50 secs
```

THIS PAGE BLANK (USPTO)

Db 301 KGMVIRCSIIPEIKGAFKCLVCHSPPLVTTVKGRVEEPTCEKPECAARNAMSLIHN 360
Qy 361 RCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTWDAVPGDRIEVTGVFKAMAVRG 420
Db 361 RCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTWDAVPGDRIEVTGVFKAMAVRG 420
Qy 421 PNQRTLALYKTYIDCVHVKKSDRGLQTEDEPMENDKENDMYAGVHESDTSEANEAKIQ 480
Db 421 PNQRTLALYKTYIDCVHVKKSDRGLQTEDEPMENDKENDMYAGVHESDTSEANEAKIQ 480
Qy 481 KLKELSKLPGIYDRLSRLAPSWELEDIKKGLLQCLFGKAKKIPSGASFRGDIINVLV 540
Db 481 KLKELSKLPGIYDRLSRLAPSWELEDIKKGLLQCLFGKAKKIPSGASFRGDIINVLV 540
Qy 541 GDPGTSKQLQYVHKIAPRGITYSGRSSAVGLTAYVTKDPETRETTLVLESGALVSDRG 600
Db 541 GDPGTSKQLQYVHKIAPRGITYSGRSSAVGLTAYVTKDPETRETTLVLESGALVSDRG 600
Qy 601 ICCIDEFDKMSDNARSMLHEVMEQTVSVAKGGIITASLNARTSVLACANPSGSRYNARLS 660
Db 601 ICCIDEFDKMSDNARSMLHEVMEQTVSVAKGGIITASLNARTSVLACANPSGSRYNARLS 660
Qy 661 VIDNIQLPPTLLSRFDLIYMLDKPDEQNDRLARHLVALHYENYEVSKQDALDIQTILTA 720
Db 661 VIDNIQLPPTLLSRFDLIYMLDKPDEQNDRLARHLVALHYENYEVSKQDALDIQTILTA 720
Qy 721 YITYARQHVHPTLSDEAAEDLINGVEMRQKNGFPGSSKKVITATPROLESIRISEALA 780
Db 721 YITYARQHVHPTLSDEAAEDLINGVEMRQKNGFPGSSKKVITATPROLESIRISEALA 780
Qy 781 RMREFSEVVEKVDAAEAVRLLDVALQOSATDHATGTDIMDLITTVGSASERIRANLLAAL 840
Db 781 RMREFSEVVEKVDAAEAVRLLDVALQOSATDHATGTDIMDLITTVGSASERIRANLLAAL 840
Qy 841 RELIADKISPGSSGLKTSQLEDIRSQSSVDVSLQDIKNALGSLQSGGFLTIVHGDIVKR 900
Db 841 RELIADKISPGSSGLKTSQLEDIRSQSSVDVSLQDIKNALGSLQSGGFLTIVHGDIVKR 900
Qy 901 V 901
Db 901 V 901

RESULT 2

US-10-768-511-8
; Sequence 8, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNER, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/10/768,511
; CURRENT FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-10-768-511-8

Query Match 100.0%; Score 4609; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 901; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MENNDALDIGAVSSPYPSQSEGVSTPLQVTSPSFDNAASPVAGRAVRQPTTSVARRRG 60
Db 1 MENNDALDIGAVSSPYPSQSEGVSTPLQVTSPSFDNAASPVAGRAVRQPTTSVARRRG 60
Qy 61 RETDSARRRRSRSLGNSVSYSSPYDAGTPTGTPVATPVYATPVGTPMGTPPSFHRGTP 120
Db 61 RETDSARRRRSRSLGNSVSYSSPYDAGTPTGTPVATPVYATPVGTPMGTPPSFHRGTP 120
Qy 121 QYKORSELGSGOKPLHRRRRSQSRPEGHRSPSRPSADGRPSESABDDTLGGEYAYVWG 180
Db 121 QYKORSELGSGOKPLHRRRRSQSRPEGHRSPSRPSADGRPSESABDDTLGGEYAYVWG 180
Qy 181 TNVNIQVLRIRAFPLHNRSSAHLNSKYIQTETVEREEDTINIDMSDIYDHDPLDY 240
Db 181 TNVNIQVLRIRAFPLHNRSSAHLNSKYIQTETVEREEDTINIDMSDIYDHDPLDY 240
Qy 241 AKIVRYPLDIIPLLDTECEVATSLPTFEKHIEARPNLKAHVHMRLEINPSDIDKLVS 300
Db 241 AKIVRYPLDIIPLLDTECEVATSLPTFEKHIEARPNLKAHVHMRLEINPSDIDKLVS 300
Qy 301 KGMVIRCSIIPEIKGAFKCLVCHSPPLVTTVKGRVEEPTCEKPECAARNAMSLIHN 360
Db 301 KGMVIRCSIIPEIKGAFKCLVCHSPPLVTTVKGRVEEPTCEKPECAARNAMSLIHN 360
Qy 361 RCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTWDAVPGDRIEVTGVFKAMAVRG 420
Db 361 RCTFANKQIVRLQETPDPAIPEGETPHTVSMCLYNTWDAVPGDRIEVTGVFKAMAVRG 420
Qy 421 PNQRTLALYKTYIDCVHVKKSDRGLQTEDEPMENDKENDMYAGVHESDTSEANEAKIQ 480
Db 421 PNQRTLALYKTYIDCVHVKKSDRGLQTEDEPMENDKENDMYAGVHESDTSEANEAKIQ 480
Qy 481 KLKELSKLPGIYDRLSRLAPSWELEDIKKGLLQCLFGKAKKIPSGASFRGDIINVLV 540
Db 481 KLKELSKLPGIYDRLSRLAPSWELEDIKKGLLQCLFGKAKKIPSGASFRGDIINVLV 540
Qy 541 GDPGTSKQLQYVHKIAPRGITYSGRSSAVGLTAYVTKDPETRETTLVLESGALVSDRG 600
Db 541 GDPGTSKQLQYVHKIAPRGITYSGRSSAVGLTAYVTKDPETRETTLVLESGALVSDRG 600
Qy 601 ICCIDEFDKMSDNARSMLHEVMEQTVSVAKGGIITASLNARTSVLACANPSGSRYNARLS 660
Db 601 ICCIDEFDKMSDNARSMLHEVMEQTVSVAKGGIITASLNARTSVLACANPSGSRYNARLS 660
Qy 661 VIDNIQLPPTLLSRFDLIYMLDKPDEQNDRLARHLVALHYENYEVSKQDALDIQTILTA 720
Db 661 VIDNIQLPPTLLSRFDLIYMLDKPDEQNDRLARHLVALHYENYEVSKQDALDIQTILTA 720
Qy 721 YITYARQHVHPTLSDEAAEDLINGVEMRQKNGFPGSSKKVITATPROLESIRISEALA 780
Db 721 YITYARQHVHPTLSDEAAEDLINGVEMRQKNGFPGSSKKVITATPROLESIRISEALA 780
Qy 781 RMREFSEVVEKVDAAEAVRLLDVALQOSATDHATGTDIMDLITTVGSASERIRANLLAAL 840
Db 781 RMREFSEVVEKVDAAEAVRLLDVALQOSATDHATGTDIMDLITTVGSASERIRANLLAAL 840
Qy 841 RELIADKISPGSSGLKTSQLEDIRSQSSVDVSLQDIKNALGSLQSGGFLTIVHGDIVKR 900
Db 841 RELIADKISPGSSGLKTSQLEDIRSQSSVDVSLQDIKNALGSLQSGGFLTIVHGDIVKR 900
Qy 901 V 901
Db 901 V 901

RESULT 3

US-10-425-114-65404
; Sequence 65404, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.


```
Db 435 LKELSKLPDIYDRLTRSLAPNIWELDDVKRGLLCQLFGNALRLSPGASFRGDIINILVG 494
Qy 542 DPGTSKSLQVYVHKIAPRGITSGSASVGLTAYTKDPETRTVLESALVLSDRGI 601
Db 495 DPGTSKSLQVYVHKIAPRGITSGSASVGLTAYTKDPETRTVLESALVLSDRGV 554
Qy 602 CCIDFDDKMSDNARSMLEHVMEOQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSV 661
Db 555 CCIDFDDKMSDNARSMLEHVMEOQTVSTAKAGIIASLNARTSVLACANPTESRYPRLSV 614
Qy 662 IDNIQLPPTLLSRFDLIYMLDKPDEQNDRLRLARHLVALHYENYSKQDALLDQTLTAY 721
Db 615 IDNIHLPTLLSRFDLIYLLDKADEQTDRLRLAKHIVSLHPENPIEELVLDLPTLVAY 674
Qy 722 ITYARQHVHPTLSDAAEDLINGYVEMQKGNFPGSSKKVITATPRQLESIMIRISEALAR 781
Db 675 ISYARKHIQPOLSDAAEBELTRGYVEMRKGNPSGSRKKVITATARQIESLIRLSEALAR 734
Qy 782 MRFSEVVEKVDAEAARVLLDVALQOSATDHATGTTIDMDLITTVGSASERIRANLLAALR 841
Db 735 MRFSEVVEQVDVVEAFRLLEVAHQOSATDHATGTTIDMDLITTVGSASERQDRDNLVAATR 794
Qy 842 ELIADKISPGSSGLKTSQLEDIRSQSDVSVLSQDIKNALGLOGEGFLTIVHGDIVKR 900
Db 795 NLVMEKMQLOGPS-VRMIELLEBEIRKQSSMEVHLDRCAUGTLMTGAVVHGDIVKVR 852

RESULT 5
US-10-425-115-283561
; Sequence 283561, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 283561
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(551)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21708C.1.pep
US-10-425-115-283561

Query Match 42.8%; Score 1972.5; DB 4; Length 551;
Best Local Similarity 66.7%; Pred. No. 38-139;
Matches 381; Conservative 85; Mismatches 82; Indels 23; Gaps 3;

Qy 331 VTVVKGRIEPTRCPECAARNAMSLIHNRCITFANKQIVRLQETPDPAIPEGETHPTVSM 390
Db 4 VNVDRGRVTEPHICQEQCATNSMTLVHNRRCFSDKQIKLQETPDEIPEGTHPTVSV 63
Qy 391 CLYNTWDAVKPGDRIEVTGVFKAMARVGPQNORTLRALYKTYIDCVHVKKSDRGLQTE 450
Db 64 LMHDKLVADGKPGDRVEITGIVRAMSIRIGPT-----KKKTDKSLHVE 107
Qy 451 DMWMDKENDMYAGVHESDTSAAANEAKIQKLKSLKPLGIYDRLSRSLAPSINWLEDIK 510
Db 108 DTMDIDNSN-----ASKSTSEDFLSDKVEKLKSLKPLDIYERLTRSLAPNIWELDDVK 161
Qy 511 KGLLQQLFGKAKIPSGASFRGDIINVLVGPDPGTSKSLQVYVHKIAPRGITSGRGS 570
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
```

```
Db 162 RGLLQQLFGGNPLKLPSCASFRGDIINLLVGPDPGTSKSLQVYVHKIAPRGITSGRGS 221
Qy 571 AVGLTAYTKDPETRTVLESALVLSDRGICCIDFDDKMSDNARSMLEHVMEOQTVSVA 630
Db 222 AVGLTAYTKDPETRTVLESALVLSDRGICCIDFDDKMSDNARSMLEHVMEOQTVSIA 281
Qy 631 KGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYMLDKPDEQND 690
Db 282 KAGIIASLNARTSVLACANPTESRYPRLSVIDNIHLAPTLLSRFDLIYLLDKADEQTD 341
Qy 691 RRLARHLVALHYENYSKQDALLDQTLTAYTIVARQHVHPTLSDAAEDLINGYVEMRQ 750
Db 342 RRLAKHIVSLHFPENLEELVLDLQTLVSYISYARKYIQPOLSDAAEBELTRGYVEMRK 401
Qy 751 KGNFPGSSKKVITATPRQLESIMIRISEALARMRSEVVEKVDAEAARVLLDVALQOSATD 810
Db 402 RGNPSGSRKKVITATARQIESLIRLSEALARMRSEVVEVDRDVEAFRLLEVAHQOSATD 461
Qy 811 HATGTTIDMDLITTVGSASERIRANLLAALRELADKISPGSSGLKTSQLEDIRSQSS 870
Db 462 HATGTTIDMDLITTVGSASERORRENVLNATNLIAEKMQLOGPS-MRMIELLEBELRKOSS 520
Qy 871 VDVSLQDIKNALGLOGEGFLTIVHGDIVKR 901
Db 521 MEIHHMELRGALGTLMTGAVVHGDIVRRV 551

RESULT 6
US-10-425-115-283560
; Sequence 283560, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 283560
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(851)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21707C.1.pep
US-10-425-115-283560

Query Match 41.0%; Score 1887.5; DB 4; Length 851;
Best Local Similarity 46.2%; Pred. No. 1.5e-132;
Matches 423; Conservative 147; Mismatches 246; Indels 99; Gaps 23;

Qy 20 SEGVTPLPQVTPSPSFONAAAPVAGRRRAVRQTPTSAVRRRGRETDSARRRSRSLGNS 79
Db 3 SDGGSGSPPPASSP-YGRPSSPLA---VTNSSPSQPTERSG-----GRRRGS----- 46
Qy 80 VYSGPYDAGTGTGTGTPVATPVATPVGTPTMGTSFHRGTQYKQSELGSGQKPLHRRR 139
Db 47 --ASPY-ASSPSLOG-----FETP-----PHPGRRTP-----SGAGA-GAPRQPRQ 83
Qy 140 RSQREPHGRSREPSADG-RPSESABDDTLGG-----EYAVVGVGNVNIIPDLVLAIR 193
Db 84 NSTGRFP--PTPTMTMTDDVPPSSAGDDETDGGGGVDATPVFVWGNISVQDVNAAIL 141
Qy 194 RFLHNYRSSAHL-----NSKYIQIIBETVRE-EDTLNIDMSDIYDHDPLYAKIV 244
Db 142 RFLRHR-DPRDAGRDPVMDGKYMRAIHRILEGGESLDVDAHDVFDHDPDLYSKWV 200
```

245 RYPLDIIPLLDTECEQVATSLPTTEKHIEARPFNLKASVHMRNLNPSDIDKLVSVKGMV 304
 201 RYPLEVLAIFDVLMDLVARIBFLPEKHITQRIYNLKSSICILNLPNSDIERVWSIKMI 260
 305 IRCSSIIPEIKGAPFKCLVCGHSPPLVTVVKGREVEPTCEKECAARNAMSLIHNCTF 364
 261 IRCSSVPELKEAVFCLVCGHSPVVDVRGRVTEPHICQRECKATNSMTLVHNRCP 320
 365 ANKQVRLQETDPAIEGETPHTVSMCLYNTVMVDAVKPGDRLEVTGVFKAMAVRVGPQR 424
 321 SDKQIILKQETDEIPEGGTPTHTVSLVMDKLVADGKPGDRVEITGIYRAMSIRIGPTQR 380
 425 TURLYKTVI-----DCVHVKKSDRGRLQTEPDPMEMDKENDMYAGYHESD---TSEAN 475
 381 TVKSIPTKIARMRFSEVVEVDVVEAFRLLEVAMQ-QSATDHATGTIDMDLMTGISAS 439
 476 E-----AKIQKLKSLKPGIYDLRLSRLSAPSISLELEIDIKGLLQCLPGKRAKIP 526
 440 ERQRNDLVAATNRLIVEKQLGG-----PSMHMIEKXAVRAQVAELNGEPSARIP 489
 527 SGASFRGIDNVLVGDPGTSKQLQYVHKIAPRGYITSGRGSSAVGLTAYVTKDPETRE 586
 490 RGAWHYDDRR--CASDPKKGGERL-----MLTAICRDGCYSSVYXVFSNSQGE 537
 587 TVLESGALVLSRIGICIDEFDMKSNARSMLHEVMEQTVSVAKGIIASINARTSVLA 646
 538 DLIDCRAPAFRLVXCAKDEPHKMSVNASRMLHEVMEQTVSIKAGIIASINARTSVLA 597
 647 CANPGSRNARLSVIDNTQLPPTLLSRFDLIYMLDKDEQNDRLARHLVALHYENYE 706
 598 CANPTESRYNPLSVIDNTLHAPTLLSRFDLIYLLDKADEQDRLAKHIVLSHPENP 657
 707 VSKQDALDQTLTAYTYARQVHPPTLSDEAAEDLINGVEMKGNPFGSSKKVITATP 766
 658 LSELEVLDTLVISYISYARKYIQPLSDEABELTRGVEMKGNPFGSSKKVITATP 717
 767 RQLESIRISEALARMFSEVVEKVDAAEAVRLDVALQOASDTHATGTIDMDLITGVS 826
 718 ROIESLIRLSEALARMFSEVVEVDVVEAFRLLEVAMQOASDTHATGTIDMDLITGVS 777
 827 ASERTIRANLALRLIADKISPGSSSGIKTSQLESLEDSQSSVDVSIQIKNALGSLQ 886
 778 ASERQRRENVAATNRLIAEKQQLGSPS--MRMTELEELRKOSSMEIHHMLRGLGTL 836
 887 GEGFLTVMGDIVKRV 901
 837 TEGAVVIHGDVNRVR 851

RESULT 7

US-10-128-714-8558
 ; Sequence 8558, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengdi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10/128,714
 ; CURRENT FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899

; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 8558
 ; LENGTH: 1023
 ; TYPE: PRF
 ; ORGANISM: Aspergillus fumigatus
 ; US-10-128-714-8558
 Query Match 40.9%; Score 1883; DB 4; Length 1023;
 Best Local Similarity 42.3%; Pred. No. 4.2e-132;
 Matches 434; Conservative 155; Mismatches 286; Indels 152; Gaps 23;
 QY 12 VSSPYPSSEGVSTPLPQVTPSPFNAASPVAGRAVRAQTPTSAVRRGRETDSARRRRS 71
 Db 1 MSSPASSRRRG-----RPKADTATSSPARSTRSQLOTSTPTPRAADSQSATPRA 51
 QY 72 RRSRLGNSVY--SSPY-----DAGTPTG-----PGTPV-----ATPVYAT 104
 Db 52 SRLRGEAAVSSSSPMFFQSSPSKADSAETPDVRMDEPSSPMRESSTWDEGDRTPRGA 111
 QY 105 PV---GTP---MGTPSFHRTPOYKORSELGSGOKPLH-----RRRSQSREP 146
 Db 112 PTMRDSSPIRYNSSSSPTRAQNRQRSDIPSSSSGLFVSSRPSIESNRAVRRSDLHSG 171
 QY 147 GH-RSPSRE-----PSADGRPSESA-----EPD-----DTLGGEYA-YVMGTNVAI 185
 Db 172 GFLSPNRRRRVFDVANGMPATDGPDSATFSNIHPDTSBAEALGSGSTRVWGTNISI 231
 QY 186 PVLRAAIRFLIN-----YRSSAHDLNSKYIQIIEETVEREE-----DTLND 228
 Db 232 QDSMAFNFLYFQTKYRLMAEGATEDETRMGDSAEERYISMLSTMRLQGLVTSINLD 291
 QY 229 MSDI--YDHDPLYAKIVRYPLDIIPLLDTECEQVATSL----- 265
 Db 292 AKNLKAYPSTLKLHQLHAYPOEIIPLMDQTVKDVNVVELAIKEMERLRAQNRQNRG 351
 QY 266 -----LPTFEKHIEAR-----PFLKASVHMRNLNPSDIDKL 297
 Db 352 LSSGPAVPSSDALSSETGRMPONEIPDLVGEVETKAFKVPFLGDLSTVNRDLDPAMDKL 411
 QY 298 VSVKGMVIRCSSIIPEIKGAPFKCLVCGHSPPLVTVVKGREVEPTCEKECAARNAMSL 357
 Db 412 VSIKGLVIRTPPIIPDMKEAFRCQVCNHHGVQ-VVIDRGKIAEPTCEPRFVCKERNMSQL 470
 QY 358 IHNRCCTFANKQIVRLQETDPAIEGETPHTVSMCLYNTVMVDAVKPGDRLEVTGVFKAMAV 417
 Db 471 IHNRCVFPADKQVVKLQETPDSIPDQGTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 530
 QY 418 RVGPNQRTLRALYKTVIDCVHVKKSDRGRLQTEPDPMEMDKENDMYAGYHESDTSAAEA 477
 Db 531 RVNPRORTQKSLFKYIYDLVHLVQKIDRKLGLDVTSTIEQELSEQAAGDAEQTRLLTAE 590
 QY 478 KIQKLKSLKPLGIYDLRLSRLSAPSISLELEIDIKGLLQCLPGKRAKIPSGAS--PRGDI 535
 Db 591 --EKIKRTATRPDLVYELLSSRLAPSIYEMDDVKKIGILLQFGTNTKTFKGGNPRVYRGI 648
 QY 536 NVLLVGDPTSKSQLLQYVHKIAPRGYITSGRGSSAVGLTAYVTKDPETRETVEESGALV 595
 Db 649 NILLCGDPTSKSQLLRYVHKIAPRGYITSGRGSSAVGLTAYVTRDPETRQMVLESGALV 708
 QY 596 LSDRGICCIDFDMKSNARSMLHEVMEQTVSVAKGIIASINARTSVLACANPGSRY 655
 Db 709 LSDGGICCIDFDMKSNARSMLHEVMEQTVSVAKGIIATLNAITSILASNPITGSR 768
 QY 656 NARLSVIDNIQPLPPTLLSRFDLIYMLDKDEQNDRLARHLVALHYEN--YEVSKQDAL 713
 Db 769 NPNLPVQNIDLPPTLLSRFDLVYLVLDVDEQEDRLAKHLVNMVLEDRPEHAARQEL 828
 QY 714 DLQTLTAYTYARQVHPPTLSDEAAEDLINGVEMKGNPFGSSKKVITATPRLQESMI 773

Db 4 ALLOQDESAMKHCSSLSRGSGGTPRRARRGDHHSFPHSGSPSLARQAQVPADLSLR 63
QY 87 ACTGCTGTPVATPVYATPVGTPMGTGPFHRTGTPYKORSELGSOCKPLHRRRRSQSREP 146
Db 64 ASSPG-----MGTDSF--GTPRV-----PGSAAPTLSAQAQOL--- 95
QY 147 GHRSRPSADGRSESABDDTIGGEYAYVWGTNVPDVLRAIRPL----- 196
Db 96 -----GADG-----DDIDGMVKFTWGTISLQESMNLFRDLRGFKPKYRAV 137
QY 197 HNYRSSADLNS-----KYIOIIBETVE-----REEDTLNIDMDIYD 234
Db 138 YNAQSRRRAESGGVAPPMTLYDNLSEAEVPLYETVYLNRLRTGETNLMDLNLIA 197
QY 235 HDP--DLYAKIVRYPLDIPLDTRCQEVAT-----SLTTFEKHI-E 274
Db 198 YRPTKGLYQOLVYQVEVPIPDQVLRDVMIELGHEELEKAKTFAEGLNSQDVESRYVK 257
QY 275 APPFNKLKASVHRELNPSSIDKLVSVKGVNIRCSIIPEIKGAPFKCLVGHSPPLVTVV 334
Db 258 VRPFGEKTVNRLNPGDGLKLVTKGLIVIRATVPIDMTTAFRCCLVQHTVQ-ADID 316
QY 335 KGRVEPTCEKPECAARNAMSLIHNRCCTFANKQIVRLOETPDATPEGETPHTVSMCLYN 394
Db 317 RGRISPERCPDRVCGSTGTMSLIHNRSFTDKQVIRLOETPDVDPDQGTPTHTVSLCYD 376
QY 395 TWVDVAKPGDRTEVGVKAMAVRVPNORTLRALYKTYIDCVHVKKSDRGRLQTEDPME 454
Db 377 ELVDLVKPGDRVITIGIFASIPVRVNPQRORSIKSYKYLYDVVHVKTNTARWGF-DPST 435
QY 455 MDKENDMAYGHESPTSEANBAKI QKLKELSKLPGIYDRLSRSLAPSWELEDIKKGLL 514
Db 436 RAGESKP-PGDLGSPVRSAAEWE-QRIETELSNHNPOLYNILASSLAPSYLEDEDVKKGIL 493
QY 515 COLFGKAKKIPSGAS-----PRGDIVLLVGDPTGSKOLLYVHKIAPRGYITSGRGS 569
Db 494 LQFGGTNKSIAKGGGGGPRYRGDIVNLVWGDPTGSKSQILQYVHKIAPRGYITSGRGS 553
QY 570 SAVGATAYVTKDPETRETVELSGALVLSRGICCIDFDMKSDNARSMLHEVMEQQTUSV 629
Db 554 SAVGATAYVTRPDSKQLVLESALVLSGGVCCIDFDMKSDATRSVLHEVMEQQTUSI 613
QY 630 AKGGIISLARTSVLACANSGSYNARLSVIDNIQLPPTLLSRPDLIYMLDKPDEQN 689
Db 614 AKAGIITLNTATSILAAANPINSRYDNLPIPIANIDLPPTLLSRPDLIYMLDKVDEYN 673
QY 690 DRLARHLVALHYENYSKQD-ALDQTLTAYITVAROHVPTLSDEAAEDLINGVEM 748
Db 674 DRKLAKHLVGLYLSDVEDQADNIIPLQTLTSYITVARSKIHPVLTEGASEALVQAYVEM 733
QY 749 RQKGNPPGSSKKVITATPQLESIMIRISEALARMRESEVVEKVDAAEAVRLLDVALQOSA 808
Db 734 RKAGMDSRTQEKRIATATTQLESIMIRIGLBAHARMLSDRVEBEDIREAVRLIKSALRESA 793
QY 809 TDHATCTIMDLITTVGSASERIRANLALARELITADKISPGSSGLKTSOLLEDIRSQ 868
Db 794 TDPLGQIDLDINTGAGQTMARADLKEVILKLVKEK---ARSQIRWAAVIDELNKQ 850
QY 869 SSVDSVLQDKNALGSLQEGFLTVHGDIVKRV 901
Db 851 SSVVPDHAQFAEIVRELEESIVKVMGERBRI 883

RESULT 10

US-10-369-493-3801

; Sequence 3801, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIORITY FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIORITY FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3801
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(960)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3801

Query Match 39.8%; Score 1835; DB 4; Length 960;

Best Local Similarity 41.2%; Pred. No. 1.6e-128;

Matches 412; Conservative 163; Mismatches 248; Indels 178; Gaps 23;

QY 24 STELPQVTPSPFD-----NAAAPVAGRAVRQTPTSVAVRRRGRETDSARRRRSRSL 76

Db 1 SPSLRQQTQTSQDADRTFRANGRSQIIGMAVWQ-----KKKKDYTECX 45

QY 77 GNSVSSSPYDAGTPTGTPVATPVYATPVGTPMGTPSFHRTGTPYKORSE-----LGSQ 131

Db 46 PHASDSSPIRYASSSGRQL-----TQSDLRSESSQLFVSSQ 84

QY 132 GKPLHRRRRSQSRECHRSR-----EPSADGRPSSESABD-----DTL 171

Db 85 RSVAGRSRRGDTGNGPLRTPAQIPRRILDDAGRVIRDA PGSDANS FVTNNPTSEADAL 144

QY 172 GGE-YAYVWGTNVPDVLRAIRFLHN-----YRSSAH-----DLNSK-YIQII 214

Db 145 GGQSGLVWGTISLDDSFSAFQFLRNFTRKRWADGADAEATIGHDPADSKPYEAL 204

QY 215 EBTVEREDTLNIDMSDI--DYHDPLDYAKIVRYPLDIPLDIT---EC-----QEVAT 263

Db 205 ENMLLGTNKLVDLRLDKSYPTLKLWHQAQHYPTLIPVMDQCVHDCWMLAQKEMAS 264

QY 264 -----SLLTPEKHE-----ARFNLK 281

Db 265 QRASQNSRTAPGASQSEBPNPSPSERSEPTTPRPAQTAAPTIEDQVSMAYVVRPWGLD 324

QY 282 ASVHRELNPSSIDKLVSVKGVNIRCSIIPEIKGAPFKCLVGHSPPLVTVV--KGRVE 339

Db 325 KITNRLDNPSSMDKLVSIKGLVIRTPVIPDMKQAFKSCVCGHS---ITVQLDRGKIR 381

QY 340 BPTCEKPECAARNAMSLIHNRCCTFANKQIVRLOETPDATPEGETPHTVSMCLYNTWDA 399

Db 382 ETECPARCAKSNMQIIHNRCADKQVILKQETPDNVPAGQTPHSVSVYVNLVDF 441

QY 400 VKPGRIEIVGVFKAMAVRVPNORTLRALYKTYIDCVHVKKSDRGRLQTEDEMD--K 457

Db 442 CKAGDRVELTIGFKVTPVRVNPMTVKSVMHTYDVVHVQVDRKRMGS-DPSTLDLAE 500

QY 458 ENDMYAGHESDTSSEANEAKI QKLKELSKLPGIYDRLSRSLAPSWELEDIKKGLLQOL 517

Db 501 EEEAHANGQMDVEVRKVSPEERIKETAAREPDYDLSRSLAPSIIYEMDDVKCKGILLQL 560

QY 518 FGGKAKKIPSGAS--PRGDIVLLVGDPTGSKQLLYVHKIAPRGYITSGRSSAVGLT 575

Db 561 FGGTNTKPKGSGPKYRGDIVLLCGDPTGSKQLLYVHRIAPRGYITSGKSSAVGLT 620

QY 576 AYVTKDPETRETVELSGALVLSDRGICCIDFDMKSDNARSMLHEVMEQQTUSVAKGII 635

Db 621 AYVTRDPESRQLVLESALVLSGGVCCIDFDMKSNSTRSVLHEVMEQQTUSVAKGII 680

QY 636 ASLNARTSVLACANSGSYNARLSVIDNIQLPPTLLSRPDLIYMLDKPDEQNDRLAR 695

Db 681 TTLNARTSILASANPTGSRYNPDLSVPQNIPLPPTLLSRPDLVYLLIDRDEKNDQRLAR 740


```
QY 110 MCTP-SFHRGTPOYKQSELGSGKPLHRRRSQSREPGHRSPPSADGRPSSESAPD 168
Db 160 EGTSPRGVGTPE-VQRPDLGSAQGLQVLDQS-----DGAADIVAS 202
QY 169 DTLGGEYAVWGTNNVNPVLAIRRFLLNY-----RSSAHLNSK-YIQIIEETVE 219
Db 203 EQSLGQKLVINGTDVNVAAKENFORFLQRFIDPLAKEENYGVIDITEPLYMQLGGINV 262
QY 220 REEDTLNIDMSDIYDHPDPLAKIVRYPLDIIPLDTECOEVATSLP--TFEKHEARP 277
Db 263 IGEQFLNVNCEHKSFQKMLYQLISYPOEVIPTFDMAVNEIFFDYPDSILHEQIQRP 322
QY 278 FNLKASVHMRRLNPSIDIKLVSGKMGVIRCSIIPEIKGAPFKCLVCGHSPPLTVVKG 337
Db 323 FNALTKNMRNLNPEDIDQLITISGWVIRTSOLIPEMQEAFTQCQCAHT-TRVEMDRGR 381
QY 338 VBEPTCKEPCCAARNANSLIHNRCYFANKQIVRLOETPDALPEGETHTVTMCLYNTWV 397
Db 382 IAPSPVCGR--CHTTHSMALIHNRSLFSDKOMIKLOESPEDMPAGQTPHTVILFAHNDLV 439
QY 398 DAVKPGDRIEVTGVFKAMAVRQPNQRTALRYKTVICVHVKSDRGRLQTEDPMEMDK 457
Db 440 DKVQPGDRVNTGTYRAVIRNPRVSNVKSVKTHIDVIHRYKTDKXRLH-----490
QY 458 ENDMVAGYHESDTSEANEAQIKLAKELSKLPGIYDRLSRSLAPSISWELEIKKGLCOL 517
Db 491 -----GLDSEABQKLFSEKRVELLAKELSKRPDIYERLASALAPSIYEHEDIKKGILLQL 544
QY 518 FGKAKKI--PSGASPRGDI NVLLVGDPTGTSKQLQYVHKIAPRGIYTSGRGSSAVGLT 575
Db 545 FGTRKDFSGTKGKFRABINILLCGDPTGTSKQLQYVYVNLVPRQYTSGRGSSAVGLT 604
QY 576 AVYTKDPETRETVELESGALVSDRGICCIDFEDKMSDNARSMLHEVMEQOYTSVAKGGII 635
Db 605 AVYMKDPETRQLVLOTGALVSDNGICCIDFEDKMNNESTRSVLHEVMEQOYTSIAKAGII 664
QY 636 ASLNARTSVLACANPSGRYNARLSVIDNIQPPPTLLSRFDLIYMLDKPDEONRRLAR 695
Db 665 CQLNARTSVLAAANPIESQWNPCKTIENIQLPHTLLSRFDLIYMLDKPDEAYDRLAH 724
QY 696 HLVALHYEYVSKQDALDLOTLTAYITVARQHVHPTLSDEAAEDLINGVEMRQKGNFP 755
Db 725 HLVALYQSEQAEEELDMVAVLKDYIYVAHSTIMPRLESEASQALIEAYVDMRK----I 780
QY 756 GSKKVITATPROLESIRISEALARMSEVVEKVDAAEAVPLLDVALQOASATHTGCT 815
Db 781 GSSRGMVAYPRQLESIRLAEAAHAKVRLSNKYEAIDVBEAKGLHREALKQSATDPRGTI 840
QY 816 IDMDLITTVGSASERIRRRANLLAALRELIADKISPGSSSGLKTSQLEDDIRSSQSVDSL 875
Db 841 VDISIITTVGMSATSRKRELAELAKKLI---LSKKTALPKYQQLFEDIRGQSDIATYK 897
QY 876 QDIKNALGSLQEGFLTVHGDIVK 899
Db 898 DMFEALRALADDDFLTVTKTVR 921
```

RESULT 13

```
US-10-032-585-7480
; Sequence 7480, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: BO, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
```

```
; SEQ ID NO 7480
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7480
```

```
Query Match 38.1%; Score 1754; DB 4; Length 910;
Best Local Similarity 42.0%; Pred. No. 1.8e-122;
Matches 397; Conservative 155; Mismatches 259; Indels 134; Gaps 20;
```

```
QY 3 NNDALDIDGAVSPYPSQSE--GVSTPLPQVTPSPFDNAASPVAGRAVRQTPTSAV---56
Db 45. SNPGSDIGNFNQSOSQQRNDISPLHYTSSAQ-----PTSDIGQP 87
QY 57 --RRRGRETDSARRRSRSRSGNSVYSPYDAGTGTGTGTPVATPVYATPVGTMGTPS 114
Db 88 DSQRSARQDVGRIMRAQRSDVTVSSP-----117
QY 115 FHRGTPOYKQSELGSGKPLHRRRSQSREPGHRSPPSADGRPSSESAPDDTLGGE 174
Db 118 -----QRS-----RRYFTQGRNGSPNLNLSSTSAQ-FSTDPAEPNDS---P 154
QY 175 YAVVGTNNVNPVLAIRRFLL--HNYRSSA-----HDLMSKYIQIIEETVERE 221
Db 155 VVVIWGTNVSIOCCSNIFRDLFSFKYRRELEQAIEPEDEHEL--YVYNQLNIIELG 212
QY 222 EDTLNDIDMSDI--YDHPDPLAKIVRYPLDIIPLDTECOEV-----ATSL 265
Db 213 LTNLNLDAKNLLSYSTRKLYYQLINYPOEIIPIMDHTIKDCLIQIINDANATTSQAQK 272
QY 266 LPTPEKHI-EARPNLK-ASVHMRRLNPSIDIKLVSGKMGVIRCSIIPEIKGAPFKCLV 323
Db 273 LDEIETNVYTIIRPVYNNLVKGIRELNPNDIDKLVSVKGLTLRSTSIIPDMKVAFPRCNA 332
QY 324 CGHSPLTVVKGREVEETPCKEPCCAARNANSLIHNRCYFANKQIVRLOETPDALPEGE 383
Db 333 CCHTVG-VEIDRGVISEPTKCPREVCQTNSWLVLIHNRSFSDKQVLIKQETPDLVDPGQ 391
QY 384 TPTVTSMCLNTMVDVAVKPGDRIEVTGVFKAMAVRQPNQRTALRYKTVICVHVKSD 443
Db 392 THSINLCVYDLVDSCRAGDVEVCGIFRSPVPRANPQRALKNLYKTYLDIVHVKKID 451
QY 444 RGLQTE-----DPMEMKNDMVAGYHESDTSEANEAQIKLAKELSKLPGIYDRLSR 497
Db 452 KRRLGQDVTTLEHAEAKDQE-----VQVRKITAEEBAKI---KEISERDDLYELAR 502
QY 498 SLAPSTWELEDIKKGLCOLFGKAKIIPSGASPRGDI NVLLVGDPTGTSKQLQYVHKI 557
Db 503 SLAPSIYEMDMVKKGLLQLFQGTNKTFTKGGRYRGDINILCGDPTSKQSLQYVHKI 562
QY 558 APRGIYTSGRGSSAVGLTAYVTKDPETRETVELESGALVSDRGICCIDFEDKMSDNARS 617
Db 563 APRGVTYTSKGSANGLTAYITFDITKQLVLESGALVSDGVGCCIDFEDKMSDATRSV 622
QY 618 LHEVMEQOYTSVAKGGIIASLNARTSVLACANPSGRYNARLSVIDNIQPLPHTLLSRFDL 677
Db 623 LHEVMEQOYTSIAKAGIITLARTSILASANPINSRYPNLPVTVGNIDLPPLLSRFDL 682
QY 678 IYMLDKPDEONDRRLARHLVALHYEN--YEYVSKQDALDLOTLTAYITVARQHVHPTLS 735
Db 683 VYLILDKVDESIDRQLARHLTOMYLEDAPETVNANSVLPVELLTLTIQAKENFPNPMTE 742
QY 736 EAAEDLINGVEMRQKGNFPGSSKKVITATPROLESIRISEALARMSEVVEKVDAAE 795
Db 743 EGNELVRSYVEMRKGEDARSEKKITATPROLESIRISEALARMSEVVEKVDAAE 802
QY 796 AVRLLDVALQOASATHTGCTIDMDLITTVGSASERIRRRANLLAALRELIADKISPGSSG 855
Db 803 AVRLIISAIDYATDPTVGTGRIDMDMIQTGTTAQRRVQEDLVSEIMKIEE-----NNNL 857
QY 856 LKTSQLEDDIRSSQSVDSLQDIKNALGSLQEGFLTVHGDIVK 900
Db 858 IRFNDLSVKLNERSSPRVNSVINEGLRRLQEGKIMETGDSHRR 902
```


Db 115 PRGIYTSRGSSAVGLTAYTKDPETGETVLES GALVLS DKGVCCVDFDKMSDNARSML 174
Qy 619 HEVMEQQT VSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLI 678
Db 175 HEVMEQQT VSIKAGIIASLNARTSVLACANPTESRYPRLSVIDNIHLAPTL LSRFDLI 234
Qy 679 YMLDKPDEQNDRLRLARHLVALHYENYEVSKQDALDLOTLTAYTYARQHVPHTLSDEAA 738
Db 235 YLILDKADEQTDRLRLAKHIVSLHFENPNLELEVLDTLQTLVSIYSYARKYIQPQLSDEAA 294
Qy 739 EDLINGVYEMRQKGNFPGSSKKVITATPRQLES MIRI SEALARMFSEVVEKVDAAEAVR 798
Db 295 BELTRGYVEMRRKRGNSPGSRKKVITATARQIESLIRLSEALARMFSEVVEVRDVVEAPR 354
Qy 799 LLDVALQOSATDHTGTTIDMDLITTVGSASERIRRANLIALRELIADKISPGSSSGLKT 858
Db 355 LLEVAMQOSATDHTGTTIDMDLITGIGASERQRRNLVATRNLI AEKMQLGGPS-MRM 413
Qy 859 SOLLEDIRSQSVDSVLSQDIKNALGSLQGEGLTVHGDIVKRV 901
Db 414 IELLELRKQSSMETHMELRGALGTMTGAVVIHGDVRRV 456

Search completed: December 7, 2005, 17:47:33
Job time : 175 secs

THIS PAGE BLANK (USP10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:28:10 ; Search time 239 Seconds
(without alignments)
2659.753 Million cell updates/sec

Title: US-10-768-511-8

Perfect score: 4609

Sequence: 1 MENNDALDIGAVSPYPSQS.....LGSLLQEGFLTVHGDIVKRV 901

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2537	55.0	911	2	Q5JKB0_ORYSA
2	2332	50.6	720	2	Q9SIV8_ARATH
3	1909	41.4	886	2	Q86IF1_DICTDI
4	1890.5	41.0	1020	2	Q4PEI7_USTLGO
5	1884	40.9	1023	2	Q4XIR4_ASPPU
6	1873	40.6	1013	2	Q7SHS5_NEUCR
7	1870	40.6	931	1	MCM4_SCHPO
8	1856.5	40.3	989	2	Q55148_CRYNE
9	1856.5	40.3	989	2	Q5K7N5_CRYNE
10	1833	39.8	933	1	CDC54_YEAST
11	1833	39.8	1020	2	Q4IQK4_GIBZE
12	1821	39.5	862	1	MCM4_MOUSE
13	1821	39.5	862	2	Q542F4_MOUSE
14	1817.5	39.4	863	2	Q6G141_XENTR
15	1817	39.4	862	2	Q8C120_MOUSE
16	1814.5	39.4	924	2	Q6FN63_CANGA
17	1811	39.3	862	2	Q921D5_MOUSE
18	1810.5	39.3	863	1	MCM4_HUMAN
19	1809.5	39.3	863	2	Q6G077_XENLA
20	1808.5	39.2	858	2	Q5XK83_XENLA
21	1807	39.2	862	2	Q9D077_MOUSE
22	1804.5	39.2	911	2	Q6BP39_DEBHA
23	1801.5	39.1	863	1	MCM4_XENLA
24	1793.5	38.9	893	2	Q6CSV7_KLULA
25	1791.5	38.9	845	2	Q6NZV2_BRARE
26	1785	38.7	888	2	Q75AE3_ASHGO
27	1784.5	38.7	858	2	Q42589_XENLA
28	1781	38.6	924	2	Q6C2V4_YARLI
29	1754	38.1	910	2	Q59M26_CANAL
30	1754	38.1	912	2	Q59M39_CANAL
31	1690.5	36.7	677	2	Q5DTS8_MOUSE

Query Match 55.0% ; Score 2537 ; DB 2 ; Length 911 ;

32 1686.5 36.6 866 1 MCM4_DROME
33 1684.5 36.5 876 2 Q7PPI5_ANOGA
34 1613 35.0 750 2 Q6PHK9_BRARE
35 1565 34.0 817 2 Q61XQ3_CAEBR
36 1562.5 33.9 823 2 Q9SXQ8_CAEEL
37 1542.5 33.5 608 2 Q5QJ32_ENTHI
38 1514.5 32.9 896 2 Q5CTW9_CRYPV
39 1503.5 32.6 894 2 Q5CGN7_CRYHO
40 1490 32.3 632 2 Q525B3_MAGGR
41 1446.5 31.4 708 2 Q8SSE5_ENCCU
42 1356.5 29.4 556 2 Q8B060_EMENI
43 1318.5 28.6 1005 2 Q81E85_PLAF7
44 1310.5 28.4 944 2 Q7RJ52_PLAYO
45 1309.5 28.4 1005 2 Q9GR06_PLAFA

ALIGNMENTS

RESULT 1
Q5JKB0_ORYSA
ID Q5JKB0_ORYSA PRELIMINARY; PRT; 911 AA.
AC Q5JKB0_ORYSA PRELIMINARY; PRT; 911 AA.
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DE Putative replication licensing factor MCM4.
GN Name=OSJNBa0051H17.26;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
RX NCBI_TaxID=39947;
RN [1]_NUCLEOTIDE SEQUENCE.
RP SASAKI T., MATSUMOTO T., YAMAMOTO K., SAKATA K., BABA T., KATAYOSE Y.,
RA WU J., NIMURA Y., CHENG Z., NAGAMURA Y., ANTONIO B.A., KANAMORI H.,
RA HOSOKAWA S., MASUKAWA M., ARIKAWA K., CHIDEN Y., HAYASHI M.,
RA OKAMOTO M., ANDO T., AOKI H., Arita K., Hamada M., Harada C.,
RA HIJISHITA S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA IKENO M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
RA KANASAWA W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA MACHITA K., MAEHARA T., Mizuno H., Mizubayashi T., Mukai Y.,
RA NAGASAKI H., NAKASHIMA M., NAKAMA Y., NAKAMICHI Y., NAKAMURA M.,
RA NAMIKI N., NEGISHI M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA SHIMOKAWA T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA WAKI K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA ZHONG H., Iwana H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA YANO M., Jiang J., Gojobori T.,
RT "The genome sequence and structure of rice chromosome 1." ;
RL Nature 420:312-316(2002).
DR EMBL; AF004232; BAD88098.1; -; Genomic_DNA.
DR GO; GO:0003634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR GO; GO:0017111; F:nucleoside-triphosphatase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006270; F:DNA replication initiation; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001208; MCM.
DR InterPro; IPR008047; MCM_4.
DR Pfam; PF00493; MCM_1.
DR PRINTS; PR01657; MCMFAMILY.
DR PRINTS; PR01660; MCMFAMILY.
DR PRODOM; PD001041; MCM; 1.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00350; MCM; 1.
DR PROSITE; PS00847; MCM_1; 1.
DR PROSITE; PS00847; MCM_1; 1.
DR PROSITE; PS00847; MCM_1; 1.
DR PROSITE; PS00847; MCM_1; 1.
SQ SEQUENCE 911 AA; 100417 MW; 94B01E5DD773C1405 CRC64;

Best Local Similarity 57.2%; Pred. NO. 4.8e-132;
Matches 514; Conservative 137; Mismatches 176; Indels 72; Gaps 16

Qy	24	STPLQVPTSPFDNAASPAGRRARRVROPTSAVRRRRGRETSDARRRRSRSLGNSVYSS	83
Db	13	NSPPSVSSPDV-RPSSLP--ATNSSFPQSGRRG-----GRRRRS-----AS	54
Qy	84	PYDAGTGTGTPVATPYATPVGTGPMGTSPFHRGTPOYKORSELGSGCKLHRRRRSQS	143
Db	55	PYPS-SPSLGG-----FETPHGRRTPSGAAARQORQ-NWTGGRFP-----96	
Qy	144	REPGRSPRRSPSADGRP--SESAPP--DTLGG-----EYAYTGNVNIPOVLR	190
Db	97	-----TPSTPMSTDVPLSSEAGDEDPETDGGGGGAGADATPVFVWGTNISQDVNA	150
Qy	191	AIRFLPHNRRSSADL-----NSKYIQIIEETVERE--EDTLNIDMSDIYDHDPPDLYA	241
Db	151	AILRFLRHRF--DPDAGRDPVMDGEGKYMRAIHRILELEGESLDVNAHDVFDHDPDLYG	209
Qy	242	KIVRYPDIIDIPLLDTECOEVATSLPTPEKHIEARPFLNKASVHMRRLNPISIDIKLVSVK	301
Db	210	KWVRYPLEVLAI FDIVLMDLVARTEPLEKHIQTRIYNLKSVCVLRNLNPSDIEKWSIK	269
Qy	302	GMVTRCSIIPEIKGAPFKVLVCGHSPPLVTVVKGRVEEPTRCSEKPECAARNAMSLIHR	361
Db	270	GMIIRCSSVIPELKEAVPRCLVCGFYSPVPMVDRGRVTEPHICQEQCKATNSMTLVHR	329
Qy	362	CTFANKQIVRLOETPDATPEGETPHTYSMCLYNTMVDVAKPGDRLEVTVGFKAMVRVGP	421
Db	330	CRFADKQIKLQETPDEIPEGGTHTYSVLMDKLVDAGKPGDRVEITGIYRAMSIRVGP	389
Qy	422	NQRTLALYKTYIDCVHVKKSDRGRLQTPEDPMWMDKENDMTAGYHESDTSAAANEAKIQ	481
Db	390	TQRTVKSIFKTYIDCLHIKKTDKSRHLVEDSMETDNP-----ANKTTEDFLDKVEK	443
Qy	482	LKELSKLPDIYDRLRSRSLAPSINWELEDIKKGLLQCLFGGKAKKIPSGASFRGIDNVLVG	541
Db	444	LKELSKLPDIYDRLTRSLAPNIWELDDVKRGLLQCLFGGNALRLPSGASFRGIDNILLVG	503
Qy	542	DPGTSKQLQYVHKIAPRGITYTSGRGSSAVGLTAYVTKDPTRETVLSEGALVSDRGI	601
Db	504	DPGTSKQLQYMHKLSPRGITYTSGRGSSAVGLTAYVTKDPTRETVLSEGALVSDRGI	563
Qy	602	CCIDFDDKMSUNAKSMHLHEVMEQQTVSAKGIIASLNARTSVLACANPSGRYNARLSV	661
Db	564	CCIDFDDKMSUNAKSMHLHEVMEQQTVSAKAGIIASLNARTSVLACANPSGRYNARLSV	623
Qy	662	IDNIQLPTLLSRFDLIYIMLDKPEDQNDRLARHLVALHYENYEVSVQDALDLQTLTAY	721
Db	624	IDNIHLPTLLSRFDLIYILDKADEQTDRLAKHIVSLHFENPNIIEVLVDLPTLVAY	683
Qy	722	ITYARQHVHTLSDEAAEDLINGYVEMRKQNPFGSSKKVITATPRQLESIMRISSEALAR	781
Db	684	ISYARKHIQPLSDSEAAELTRGYVEMRKGNPSGRSKKVTATATARQTESIRLSEALAR	743
Qy	782	MRPSEVKEVDAAAVRLLDVALQOSATDHATGTTIDMDLITTVGSASRIRRRANLAALR	841
Db	744	MRPSEVKEVDVVEAFRLLEVAMQOSATDHATGTTIDMDLITTVGSASRIRRRANLAALR	803
Qy	842	ELIADKISPGSSSGKLTSQLLEDIRSSQSDVSVLQDIKNALGSGQEGFTLVHGDIVKR	900
Db	804	NLVMEKMQLGGPS-VRMIELLEERIKQSSMEVHLHDLRGALGTLMTEGAVVHGDIVKR	861

RESULT 2

Q9Siv8 ARATH

ID Q9Siv8 ARATH PRELIMINARY; PRT; 720 AA.

AC Q9Siv8;

DT 01-MAY-2000 (TremBLrel. 13, Created)

DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)

DE Putative CDC21 protein.

GN Name=At2g16440;


```
QY 629 VAKGGIIASLNARTSVLACANPSGRYNARLSVIDNIQPPPTLSRFDLIYIMLDPQBEQ 688
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 469 IAKAGIIASLNARTSVLACANPSGRYNARLSVIDNIQPPPTLSRFDLIYIMLDPQBEQ 528
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 689 NDRRLARHLVALHYENYVSKODALDLOLTATYITVAROHVHPTLSDEAAEDLINGVEM 748
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 529 TDRRLAKHLVALHYENYVSKODALDLOLTATYITVAROHVHPTLSDEAAEDLINGVEM 588
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 749 RQKGNFPGSSKKVITATPQRLSEMRIRSEALARMFSEVVEKVDAAEAVRLDVALQQSA 808
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 589 RQAGNFAGSSKKVITATPQRLSEMRIRSEALARMFSEVVEKVDAAEAVRLDVALQQSA 648
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 809 TDHATGTTMDLITTVGSASERIRANLLAARELIADKISPGSSGLTKTSOLLEDIRSQ 868
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 649 TDHATGTTMDLITTVGSASERIRANLLAARELIADKISPGSSGLTKTSOLLEDIRSQ 704
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 869 S---SVDVSLQIDKN 880
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 705 GGNINTEIHLHDVSH 719
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
Q86IF1_DICDI PRELIMINARY; PRT; 886 AA.
AC Q86IF1;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)
DE Similar to essential for initiation of DNA replication; homolog of S.
DE pombe CDC21; Cdc54p [Saccharomyces cerevisiae] (Hypothetical
DE protein).
GN ORFNames=DOB0167140;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910; DOI=10.1038/nature00847;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J.A.,
RA Bankier A.T., Dear P.H., Lehmann R., Baumgart C., Farra G.,
RA Abril J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A.,
RA Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugchang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero P.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Pey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren C., Davis P.,
RA Farbrother A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Karbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulaged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA Chisholm R.I., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
```

```
RL Nature 0:0-0(2005).
DR EMBL; AC116551; AAC02157.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR GO; GO:0006270; F:DNA replication initiation; IEA.
DR InterPro; IPR001208; MCM.
DR InterPro; IPR008047; MCM_4.
DR Pfam; PF00493; MCM; 1.
DR PRINTS; PR01657; MCMFAMILY.
DR PRINTS; PR01660; MCMPROTEIN4.
DR PRODOM; PD001041; MCM; 1.
DR SMART; SM00350; MCM; 1.
DR PROSITE; PS00847; MCM_1; 1.
DR PROSITE; PS0051; MCM_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 886 AA; 100648 MW; DD545AA3ED9421DF CRC64;

Query Match 41.4%; Score 1909; DB 2; Length 886;
Best Local Similarity 49.8%; Pred. No. 3.2e-97;
Matches 397; Conservative 128; Mismatches 174; Indels 98; Gaps 14;

QY 179 WGTWVNIPIVLRAIRRFLLNY-----RSSAHLNSK----- 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 103 WGTTVKVEDVKERFRKFFIFNPTPKERKRFNIDNNNNNNNNNNSEEMDDDDYYDE 162
-----YIOIIEETVEREDTLNIDMSDIYDHPDLYAKIVRYPLDIPL 254
QY 210 -----YIOIIEETVEREDTLNIDMSDIYDHPDLYAKIVRYPLDIPL 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 EQDKLIKPEKVQYLYRDLKVLNETKKCHLNIMKYLQVDFMELYLOWVSPFENMIPLL 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 255 DTECOEVATSLPTTFEKH-----IEARPNLKASVHREINLPSDIDK 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 223 DEEINLIYRELFPITPRNGDDDDDDDEDEDEYRIELHPFNLLARKTPMRDLNPSDIDK 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 LVSVKGMVIRCSIIPEIKGAFKCLVCG--HSPLVTVVVGRVEEPTCEKPCAARN 353
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 283 IISISGLIIRSSIIPEIKQAFMCAVCEATPHA----NVEKGKIQEPSECS--NCKSKQ 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 354 AMSLIHNRCTFANKOIVRLQETPDALPEGETPHTTYSMCLYNTWDAVKPDGRIEVTGVFK 413
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 337 SUSIIHNRCLFGDKQYIKLQETPDALPEGETPHTTVALFAYGLIDIATKAGDRVELTGVFK 396
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 414 ANAVRGNQRTLRALYKTYIDCVHVKKSDRGLQTEDEPMEMD-----KENDMYA 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 397 ASPMRAGSN-RSLRSYKTYIDILHKRTDKGYDDDDDDDDNTGGTGTGKETNEML 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 464 GYHESDTSSEANEAKIQKLKELSKLPGIYDRLSRSLAPSIMWELEDIKKGLLQCLFGKAK 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 DFDLDFDLSEE-KEREIIELSKKPDYIDIVTKSIAPNIWELEDIKKGLLQCLFGSKK 514
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 524 KIPS-CASPRGDIINVLLGDPGTSKQLQYVHKIAPRGITSGRGSASVGLTAYVTKDP 582
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 515 SYQDYGGKFRGDIINLLCGDPGTSKQLSYVHKIAPRGITSGRGSASVGLTAYVTKDP 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 583 ETRETVLESGLVLDGRGCCIDEDFKMSDNARSLHVEYMEQOTVSVAKGIIITASNART 642
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 575 DTKETVLESGLVLDGRGCCIDEDFKMNDQTFSLHVEYMEQOTVSIAKAGIIITASNART 634
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 643 SVLACANPSGRYNARLSVIDNIQPPPTLSRFDLIYIMLDPQBEQNDRLARHLVALHY 702
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 635 SILASANPSGRYMPKLSVVENIQLPPTLSRFDLIYILVDKANERSDRLARHLVSMYV 694
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 703 ENYEVSKQDALDLOLTATYITVAROHVHPTLSDEAAEDLINGVEMRQGNFPGSSKKVI 762
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 695 DETPVS-HFTIPKETLTNTYIQAARKHNPKLTDSSAKCLVQGYLENRSM----GSSKTKI 749
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 763 TATPROLESIRIRSEALARMFSEVVEKVDAAEAVRLDVALQQSDTATGTTIDMDLIT 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 750 SATPROLESIRIRAEAHARIRSEFVEPLDVEAIRLIKVALQQAIDENGTTIDMDLIT 809
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 823 TGVVASERIRANLLAARELIADKISPGSSGLTKTSOLLEDIRSQSSVDV--SLQIDKN 880
```

Db 810 TGRSASSRRAITRLKSHKIQKIGKK-----HLTDQLLKLLTQNGVOTIQOIBEIKE 862

QY 881 ALGSLQGGEGFLTVHGGDI 897

Db 863 ALRQLQDEIIQSSGGI 879

RESULT 4

Q4PE17 USTMA PRELIMINARY; PRT; 1020 AA.

AC Q4PE17_ USTMA PRELIMINARY;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein.

GN OFNames=UW01646.1;

OS Ustilago maydis S21.

OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;

OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.

OX NCBI_TaxID=237631;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=521;

RA Birren B., Nubaum C., Abebe A., Abouelleil A., Adekoya E.,

RA Ait-zahra M., Allen T., Allen T., An P., Anderson M., Anderson S.,

RA Arachchi H., Armbruster J., Bachantang P., Baldwin J., Barry A.,

RA Bayul T., Blitchstejn B., Bloom T., Blye J., Boguelavskiy L.,

RA Borowsky M., Boukhgalter B., Brunack A., Butler J., Calixte N.,

RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,

RA Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,

RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,

RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,

RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,

RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,

RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,

RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,

RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,

RA Jaffe D., Jones C., Kamal M., Kamat A., Kamvaselis M., Karlsson E.,

RA Kells C., Kieu A., Kiener P., Kodira C., Kulbokas E., Labutti K.,

RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,

RA Lindblad-toh K., Liu X., Lokytsang T., Lokytsang Y., Lucien O.,

RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,

RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,

RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,

RA Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Menga V., Moru K.,

RA Mozes J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,

RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari N., Norbu C.,

RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotsho B.,

RA O'neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pigani B.,

RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,

RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,

RA Rutman M., Schupbach R., Seaman C., Settippalli S., Sharpe T.,

RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,

RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,

RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,

RA Tenzing P., Teefaye S., Theodore J., Thoulutsang Y., Topham K.,

RA Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,

RA Venkataxam V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,

RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,

RA Yang S., Yang X., Yeager S., Yee B., Young G., Zainoun J., Zembeck L.,

RA Zimmer A., Zody M., Lander E.

RT "The genome sequence of Ustilago maydis."

RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

CC -1- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC preliminary data.

DR EMBL: AACP0100063; EAK82427.1; -; Genomic_DNA.

KW Hypothetical protein.

SQ SEQUENCE 1020 AA; 111539 MW; 865EBB3D96E87470 CRC64;

Query Match 41.0%; Score 1890.5; DB 2; Length 1020;

Best Local Similarity 42.2%; Pred No. 4.2e-96;

Q4X1R4 ASPFU PRELIMINARY; PRT; 1023 AA.
AC Q4X1R4;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE DNA replication licensing factor Mcm4, putative.
GN ORFNames=Afu2g09060;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Sordariales; Trichocomaceae; mitospore Trichocomaceae; Aspergillus.
OX NCBI_TaxID=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Nierman W., Pain A., Anderson M.J., Mortman J., Kim H. Stanley.,
RA Arroyo J., Beriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,
RA Haas H., Harris D., Horuchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
RA Majoros W.H., May G.S., Miller B.L., Mohammad Y., Molina M., Monod M.,
RA Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
RA Penalta M.A., Perteau M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Roming C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B., Denning D.W.;
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
RT Aspergillus fumigatus";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; RAHF0100001; EAL91201.1; -; Genomic DNA.
SQ SEQUENCE 1023 AA; 113033 MW; 33975718956DDFB6 CRC64;

Query Match 40.9%; Score 1884; DB 2; Length 1023;
Best Local Similarity 42.4%; Pred. No. 9.7e-96;
Matches 435; Conservative 154; Mismatches 286; Indels 152; Gaps 23;

QY 12 VSSPTPSQEGVSTPLPQVTSFDMASPVAGRAVRQTPTSVAVRRGRETSARRRS 71
Db 1 MSSPASRRRG-----RPAKOTATSSPARSTRSRLQTSPTTPRAADEQSQA TPRA 51

QY 72 RRSLSGNSVY--SSPY-----DAGTGTCT-----PGTPV-----ATPVYAT 104
Db 52 SRLRLGEAAVPSSPMFSSPKADSSAETPDVRMDEPSSPMRESSTWDEGRTPRGNA 111

QY 105 PV---GTP---MGTPTFHRGTPOYKORSELGSGKPLH-----RRRSQSREP 146
Db 112 PTWRDSPTRYMSSSPTTAQNRQSRSDIPSSSGLFVSSRPSIBSNVRSRSLHSG 171

QY 147 GH-RSPSRE-----PSADGRPSGA-----EPD-----DTLGEYA-YVMGNVNI 185
Db 172 GFLSSPNRRRVFVDANGMPATDGDPRSDATPSNIHPDTSSEALGSGSTRVWGTNISI 231

QY 186 PDVLRAIRPLHN-----YSSAHLNSKYIQTIEETVEREE-----DTLND 228
Db 232 QDSMSAFKPNFLYNFQTKYRLWAEGATEDETRIMGDSABEREYISMLSTWRQLGVTSLND 291

QY 229 MSDI---YDHPDPLYKIVRYPLDIIPLDTECOEATSL----- 265
Db 292 AKMLKAYPTLKLWHLQYHAYPQEIIFLMDQTVKDVVWVLAKEMERLRAQNQNHNRG 351

QY 266 -----LPTFEKHIEAR-----PFLNKASVHMRELNPSPIDKL 297
Db 352 LSSGPAVPSSDALSGETGRMPQNEIPDLVGEVETKAFKVLFPGLDSTVNRDLDPADMDKL 411

QY 298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPLVTVVVKGRVEEPTRCBKEPCAAARNASL 357
Db 412 VSIKGLVIRTTPIIPDMKEAFRCQVCHGVQ-VDIRGKVAEPTECPRPVCKEENSML 470

QY 358 IHNCTFRANKOIVRLQETPDALPSGETPHTVSMCLNTYMDAVKPGDRIEVTGVKAMAV 417
Db 471 IHNRCVPADKQVILQETPDSIPDQTPHSVSLCVYDELVDVCKAGDRVETGIFCNPV 530

QY 418 RVGNQRTLRALYKTYIDCVHVKKSDRGLQOTDEPMENDKENDMYAGYHESDTSBAANE 477
Db 531 RVNPRQRTQKSLFKTYIDVLHVQKIDRKLGIDVSTIHOELSEQAAGDAEQTRRLTABEE 590

QY 478 KIQKLEKSLKPGIYDRLSRSLAPSINWELDIKGLLQCLPGKAKKIPSGAS--PRGDI 535
Db 591 --EKIKRTATRPDLPELLESLAPSIEYEMDDVKKGILLQLFGGTNKTQKGNPNRYGDI 648

QY 536 NVLLVGDPGTSKQLQVVKIAPRGIVTSGRSSAVGLTAVVTKDPETRETVLSSGALV 595
Db 649 NILCGDPSKSKQLRIVVKIAPRGIVTSGRSSAVGLTAVVTKDPETRETVLSSGALV 708

QY 596 LSDRGICCIDFDMKSDNARSMLHEVMEQQTVSVAKGIIASLNARTSVLACANPSGSR 655
Db 709 LSDGICCIDFDMKSDNARSMLHEVMEQQTVSVAKGIIASLNARTSVLACANPSGSR 768

QY 656 NARLSVIDNIQLPPTLLSRFDLIYMLDKPQNDRRRLARHLVALHYEN--YEVSKQDAL 713
Db 769 NPNLPVQNIIDLPTLLSRFDLIYMLDKPQNDRRRLARHLVALHYEN--YEVSKQDAL 828

QY 714 DLOTTTAYTARQVHPTLSDEAAEDLINGVEMKQNGFPGSSKVVITATPROLESMI 773
Db 829 PIEFTAYTAYTAKTVKPVLTTPAAGKALSADAYVNRKLGDDIRSDRRITATPROLESMI 888

QY 774 RISEALARMFSEVVEKVDAAEAVRLDVALQOSATDHATGTIDMDLITTVSASERIRR 833
Db 889 RLSEAHARMLRSLPEVTADDVEAVRLIRSAIKQANTDSTGLIDMSLLTEGTSASERIR 948

QY 834 ANLLAALRELITADKISPGSSGLKTSQLEDIRSSQSVDSVLSQDIKNALGSLQEGFLTV 893
Db 949 EALKREALLSVVDDLCGGGAA--RWAEEVFRILSENSSIEVDGAFADAVRALEAGAVSV 1006

QY 894 HGDIYVR 900
Db 1007 VGEARR 1013

RESULT 6
Q7SHS5 NEUCR PRELIMINARY; PRT; 1013 AA.
AC Q7SHS5;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU02539.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariales; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Ekins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Seltrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothle G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysysselis M., Mauceli E., Bielek C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Omani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,

```
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannheim G., Ebbola D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nuebaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL: AABX01000002; EAA36434.1; -; Genomic_DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR GO: GO:0006270; P:DNA replication initiation; IEA.
DR InterPro: IPR001208; MCM_1.
DR InterPro: IPR008047; MCM_4.
DR Pfam: PF00493; MCM; 1.
DR PRINTS: PR01657; MCMFAMILY.
DR PRINTS: PR01660; MCMPROTEIN4.
DR ProDom: PD01041; MCM; 1.
DR PROSITE: PS00847; MCM_1; 1.
DR PROSITE: PS50051; MCM_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1013 AA; 111576 MW; 1980D7DB790AFF3E CRC64;

Query Match 40.6%; Score 1873; DB 2; Length 1013;
Best Local Similarity 41.9%; Pred. No. 3.9e-95;
Matches 423; Conservative 168; Mismatches 257; Indels 161; Gaps 26;

QY 18 SQSGVSTPLPOVT-----SPSPDAAAPVAGRRV-----RQ-----TPTS 54
DB 23 SSQAGPSSATPQTASQLASPLYPESSPANGAAPVSSPLQMSNTQSTAHQGNAPSS 82
QY 55 AVRREGR---ETDSARRRRSRSGNSVSPYDAGTGTGTPVATPVVATPVGTGPMG 111
DB 83 PLRQQTQTSQADRTFRANGRSQITGD---SPIRYASSSPGRQL----- 125
QY 112 TFSFRGTPTQYKQSE-----LGSQKPLHRRRSQSRPQHRSPSR----- 153
DB 126 -----TQQSLRSESSQLFVSSQSRVAGRRSRRGDDINGDPLRTPAQIPRIILDDAGRV 178
QY 154 ---EPSADGRPSESAPD-----DTLGGGE-YAVVGTNTVNIPIVLAIRRLHN-----YRS 201
DB 179 IRDAPGSDANSFVNTNPTSEADALGGQSGGLVWGTTISLSDSFSAFQDFLNFTRKTRM 238
QY 202 SAH-----DLNKK-YIQIIEETVEREEDTLNIDMSDI--YDHDPOLYAKIVRYP 247
DB 239 WADGADAEATIGHPDADSKPYWEALENMLLLGTNKLKLDRLKSYPRTLKLWHQQAQHP 298
QY 248 LDIIPLLOT---EC-----QEVAT----- 263
DB 299 TEIIPVMDQCVHDCMELAQEMASORASQNSRTAPGASQSQSEPNFPPSSERSEEPPTPRP 358
QY 264 ---SLIPLTEKHTE-----ARPNLKASVHMRNLNPSDIDKLVSVMGVRICSSIIPEIKG 316
DB 359 ACTAAPTIEDQVSQYAVVVRPGLDKITNLRDLNLSMDKLVSIGLIVRTPTVLPDWD 418
QY 317 AFFKCLVCHSPPLTVV--KGRVEEPTRCPEKCAARNAMSLIHNRCRTFANKQIVRLQE 374
DB 419 AFFKSCVCGHS---ITVQLDRGKIREPTECPARCASKNSMQIHNRCAFEDQVILKQE 475
QY 375 TPDATPEGETHTVSMCLYNTMWDVAKGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYI 434
DB 476 TPDNVPAGQTPHSVSCVYVNELVDFCKAGDRVELTGIFKVTVPVRVYPMRTVKSVMHTYV 535
QY 435 DCVHVKKSDRGRLOTEDPMEMD--KENDMYAGYHESDTSSEANEAKIKLKLKSLPGIY 492
DB 536 DVVHVQKVDKRMKGS-DRSTLDLAEERAHANGQSMDEVKRVSPDEERIKETARPDIY 594
QY 493 DRLSLSLAPSIWELEDIKKGLCOLFGGKAKKIPSGAS--FRGDIINVLVGDGPTGSKSQL 550
DB 595 DLLSRSLAPSIYEMDVKKIGILLQLFGGNTKNTFEKGGSPKVRGDIINVLVCGDPSTSKSQL 654
```

```
QY 551 LOYVHKIAPRGITYTSGSGSAVGLTAYVTQDPETRETTVLESGALVLSDRGCCIDFQDM 610
DB 655 LSYVHTAPRGVITSGSGSAVGLTAYVTRDPESRQLVLESGALVSDGGVCCIDFQDM 714
QY 611 SDNARSMHLVMEQQTYSVAKGGIITLARTSARTSVLACANPGSGRYNARLSVIDNIQLPPT 670
DB 715 NESTRSLVHVEQQTYSVAKGGIITLARTSARTSVLACANPGSGRYNARLSVIDNIQLPPT 774
QY 671 LLSRPDLIYMLDXPDQNDRLARHLVALHYENYVSKQ---DALDLOTLTAYTYARQ 727
DB 775 LLSRPDLIYMLDXPDQNDRLARHLVALHYENYVSKQ---DALDLOTLTAYTYARQ 834
QY 728 HVHPTLDEAAEDLINGVEMROKGNFPGSSKKVITATPROLESIMIRISEALARMFSEV 787
DB 835 HIHPALTPFAGRELVDAYVEMRKLQGVRAAEKKITATPROLESIMIRISEALARMFSEV 894
QY 788 VEKVDAEAVRLDLVALQOSATDHATGTIDMDLITTVGSASERIRRALALRELIADK 847
DB 895 VTRDDVREAVRLIKSALKTAATD--SQGRIDMSLLTGTSSAAERQKADKMDKDAVIRLLDEM 953
QY 848 ISPGSSGLKTSQLEDIRSOSVVDVSLQDINKALGSLQGGFGLTVHGD 896
DB 954 TSGGV--VRYSEVARRLGEGAGVQVPEPAEVRALMEGAVMTGE 1000

RESULT 7
MCM4_SCHPO STANDARD; PRT; 931 AA.
AC P29458; Q9P7K4; Q9USM0;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE DNA replication licensing factor mcm4 (Minichromosome maintenance
DE protein 4) (Cell division control protein 21).
GN Name=mcm4; Synonyms=cdc21; ORFNames=SPCC16A11.17, SPCC24B10.01;
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP NUCLEOTIDE SEQUENCE.
EX MEDLINE=93087163; PubMed=1454522;
RA Coxon A., Maundrell K., Kearsey S.E.;
RT "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT early step of chromosome replication.";
RL Nucleic Acids Res. 20:5571-5577(1992).
RN [2]
RP SEQUENCE REVISION TO C-TERMINUS.
RA Kearsey S.E.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=972;
EX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M.H., Lyne R., Stewart A.,
RA Sgouros J.G., Peat N., Hayles J., Baker S.G., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D.E., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K.D., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K.L., Murphy L.D., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K.M., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M.N., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J.R., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weitjens I., Vanstreels G., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Duesterhoeft A., Fritz C., Holzer E., Moestl D.,
RA Hilbert H., Borzym K., Langer I., Beck A., Lechner H., Reinhardt R.,
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
```

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey P., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armatrong J., Foreburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Barrell B.G., Nurse P.,
RL "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [4]
RP SUBUNIT.
RC STRAIN=SP011;
RX MEDLINE=21518581; PubMed=11606526;
RA Liang D.T., Forsburg S.L.;
RA "Characterization of Schizosaccharomyces pombe mcm7(+) and cdc23(+)
RT (MCM10) and interactions with replication checkpoints.";
RL Genetics 159:471-486(2001).
CC -!- FUNCTION: Required for S phase execution.
CC -!- SUBUNIT: Heterohexamer. The heterodimers of mcm4/mcm6 and
CC mcm3/mcm5 interact with mcm2 and mcm7.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the MCM family.
CC -!- SIMILARITY: Contains 1 MCM domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; X58934; CA41628.1; -; Genomic DNA.
DR EMBL; AL109597; CAB53089.1; -; Genomic DNA.
DR EMBL; AL157991; CAB76210.1; -; Genomic DNA.
DR PIR; S26640; S26640.
DR GenDB; Spombe; SPCC16A11.17; -.
DR GO; GO:0042555; C:MCM complex; IDA.
DR GO; GO:0006270; P:DNA replication initiation; IDA.
DR InterPro; IPR001208; MCM.
DR InterPro; IPR008047; MCM 4.
DR InterPro; IPR012335; Thioresdoxin fold.
DR Pfam; PF00493; MCM; 1.
DR PRINTS; PRO1657; MCMFAMILY.
DR PRINTS; PRO1660; MCMPROTEIN.
DR ProDom; PD001041; MCM; 1.
DR SMART; SM00350; MCM; 1.
DR PROSITE; PS00847; MCM 1; 1.
DR PROSITE; PS50051; MCM 2; 1.
KW ATP-binding; Complete proteome; DNA replication; DNA-binding;
KW Nuclear protein; Nucleotide-binding; Transcription;
KW Transcription regulation.
FT DOMAIN 493 702 MCM.
FT NP_BIND 545 552 ATP (potential).
FT CONFLICT 872 891 Missing (in Ref. 3).
SQ SEQUENCE 931 AA; 103728 MW; BCA9B045FC62811D CRC64;
Query Match 40.6%; Score 1870; DB 1; Length 931;
Best Local Similarity 44.0%; Pred. No. 5.1e-95;
Matches 408; Conservative 163; Mismatches 238; Indels 118; Gaps 21;
QY 38 AASPVAGRAVAVQTPTSAVRRGRETDSARRRRSRISGNSVYSSP--YDAGTPG-----91
DB 2 SSSQSGRANELRTP-----GRANSSSR-----EAVDSSPLFPFPASSPGSTRL 44
QY 92 -TPGTVPATPVATVGTPTMGTP-----SPHRGTPQYKORSEL 128
DB 45 TTPRTTARTPLASSPLLFSSSPGPNIPQSSRSHLLSQRNDLFLDSSQRTPRSTRGDI 104
QY 129 GS--CGKPLHRRRSQSRBGRHPSR--EPSADGRPSESAEP-----DITLGGYAYVW 179
DB 105 HSSVQVMSSTRREVDPQRPVGTPTSLFLPSGSDALTFSQHPSSVADTV-----RVIV 160
QY 180 GTNVNIPDVLRATRRRLTHNYS-----SAHDLNSKYIQIIEETVERBEDTL 225
DB 161 GTNVSIQESIASPRGFLRGFKKYPYRNELMPPDAEQL--VYIEALRNRMIGLEIL 218

QY 226 NIDMSDIYDHP--DLVAKIVRYPLDIPLDTECQEVATSLLPTE-----KHI 273
DB 219 NLDVQDLKHPPTTKLYHQLYSYQEIPIIMQDTIKDVMLDLGTNPDPVLDIELKIY 278
QY 274 EARPENLXASVHMBELNPSDIDKLVSVKGVIRGSSIIPEIKGAPFKCLVCGHSPPLVTV 333
DB 279 KIRFPNLEKCNMRDLNPGDIDKLISIKGLVLRCTPVPIDMKQAFRCVSGHC-VYVEI 337
QY 334 VKGRVEEPTRCPECAARNAMSLIHNRCITFANKQIVRLQETPDPAIPEGTHPTVSMCLY 393
DB 338 DRGRIAEPIKCPREVCGATNAMQLIHNSEFADKQVIKQLQETPDVVPDQTPSHSVLCVY 397
QY 394 NTWDAVAVKGRDRIEVTGVFKAMAVRVGNQRTLRALYTYIDCVHVKSDDRGRLQTEPM 453
DB 398 DELVDSARAGDRIEVTGIFRCVVPVRLNPRMRTVKSFLTYVVDVHVKQDKERLGT-DPS 456
QY 454 EMDKENDMYAGVHESDTEAA-----NEAKIQKLKELSLPGYIYDLSRSLAPSI 503
DB 457 TL-----ESDIAEDAALQIDEVRKISDEVEKIQQVSKRDDIYDILSRSLAPSI 505
QY 504 WELEDIKKGLLQCLFPGGKAKKIPSGAS--FRGDIVLLVGDPTGTSKSQLQVYVHKIAPRG 561
DB 506 YEMDDVKGLLLQLFEGGINKSFHKGASPRYRGDINILMCGDPSTSKSQLKIVHKIAPRG 565
QY 562 IYTSGRGSAVGLTAYVTKPRTETVLESALVLSDRGICCIDBFDKMSDNARSMHVEV 621
DB 566 VYTSKGSSAVGLTAYITRDQTKQLVLESALVLSGDCICCIDBFDKMSDATRSILHEV 625
QY 622 MEQQTYSVAKGGIIASLNARTSVLACANPSSRYNARLVSDINICLPPTLLSRFDLIYLM 681
DB 626 MEQQTVTYAKAGIITLNARTSVLACANPSSRYNARLVSDINICLPPTLLSRFDLIYLI 685
QY 682 LDKPDEQNDRLARHLVALHYEN--YEVSKQDALDQTLITAYITVARQHVHTLSDEAAE 739
DB 686 LDRVDETLDKLANHIVSMYMETTEHATDMEVSPVEFLTSYITARNINPVISEAAK 745
QY 740 DLINGYVEMRQKGNPFGSSKKVITATPQLESIRISEALAMRPFSEVVEKVDAAEAVRL 799
DB 746 ELVNAYVGMKLGEDVRASEKRIATATTQLESIRLSEAHAKMLRNVEVGVDLAEARL 805
QY 800 LDVALQOATDHATGTIDMDLITTVGSASERRRANLLAALRELIADKISPGSSSLKTS 859
DB 806 IKTAIKYATDPATKISLDLIY--VNERETLVPMVKELANLSNLTVGKGT--MLVS 861
QY 860 QLEDIRSQSSVDVSLQIKNALGSLQ 886
DB 862 QLTFRFSQSTRLDASFEACLGAFR 888
RESULT 8
Q55148 CRYNE
ID Q55148_CRYNE PRELIMINARY; PRT; 989 AA.
AC Q55148;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=CNBM1700;
OS Cryptococcus neoformans var. neoformans B-3501A.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=283643;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B-3501A;
RA Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,
RA Wickes B.L., Fu J., Davis R.W.;
RT "Cryptococcus neoformans serotype D sequencing."
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

[illegible]

RESULT 12

MCM4_MOUSE	STANDARD;	PRT;	862 AA.
ID MCM4_MOUSE			
AC P49717; O89056;			
DT 01-OCT-1996 (Rel. 34, Created)			
DT 01-OCT-1996 (Rel. 34, Last sequence update)			
DT 13-SEP-2005 (Rel. 48, Last annotation update)			
DE DNA replication licensing factor MCM4 (CDC21 homolog) (P1-CDC21).			
GN Names=Mcm4; Synonyms=Cdc21, Mcm4;			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;			
OC Muroidae; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
[1]			
RN NUCLEOTIDE SEQUENCE.			
RP MEDLINE=95334361; PubMed=7610039;			
RX Kimura H., Takizawa N., Nozaki N., Sugimoto K.;			
RA "Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs and			
RT characterization of the products: physical interaction between			
RT P1(MCM3) and CDC46 proteins."			
RL Nucleic Acids Res. 23:2097-2104(1995).			
[2]			
RN NUCLEOTIDE SEQUENCE OF 503-602.			
RP STRAIN=BALB/c; TISSUE=Spleen;			
RX MEDLINE=99012997; PubMed=9796653; DOI=10.1016/S0161-5890(98)00031-5;			
RA Chu C.C., Paul W.E.;			
RT "Expressed genes in interleukin-4 treated B cells identified by cDNA			
RT representational difference analysis.";			
RL Mol. Immunol. 35:487-502(1998).			
CC -1- FUNCTION: Involved in the control of DNA replication.			
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).			
CC -1- SIMILARITY: Belongs to the MCM family.			
CC -1- SIMILARITY: Contains 1 MCM domain.			

CC This Swiss-Prot entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use as long as its content is in no way modified and this statement is not			
CC removed.			

DR ENBL; D26089; BAA05082.1; -; mRNA.			
DR ENBL; U89402; AAC36509.1; -; mRNA.			
DR PIR; S56766; S56766.			
DR Ensembl; ENSMUSG0000022673; Mus musculus.			
DR MGI; MGI.103199; Mcm4.			
DR InterPro; IPR003593; AAA_ATPase.			
DR InterPro; IPR001208; MCM.			
DR InterPro; IPR008047; MCM_4.			
DR Pfam; PF00493; MCM; 1.			
DR PRINTS; PR01657; MCMFAMILY.			
DR PRINTS; PR01660; MCMPROTEIN4.			
DR Prodom; PD001041; MCM; 1.			
DR SMART; SM00382; AAA; 1.			
DR SMART; SM00350; MCM; 1.			
DR PROSITE; PS00847; MCM 1; 1.			
DR PROSITE; PS50051; MCM 2; 1.			
KW ATP-binding; DNA replication; Nuclear protein;			
NW nucleotide-binding; Transcription; Transcription regulation.			
KW DOMAIN 457 666 MCM.			
FT NP BIND 509 516 ATP (Potential).			
FT FT CONFLICT 530 530 Q -> R (in Ref. 2).			
FT FT CONFLICT 572 572 I -> T (in Ref. 2).			
SQ SEQUENCE - 862 AA; 96736 MW; 516ACC1A3C6FB16E CRC64;			
Query Match 39.5%; Score 1821; DB 1; Length 862;			
Best Local Similarity 43.7%; Pred. No. 2.4e-92;			
Matches 401; Conservative 154; Mismatches 258; Indels 104; Gaps 18;			
QY 30 VTSPSFDAASPVAGRAVRQ---TPTSAVRRGRGTDSAPRRSRSLG----- 77			

THIS PAGE BLANK (USP 201)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 09:08:24 ; Search time 7664 Seconds
(without alignments)
6682.666 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALDCAVSSPYSPQS.....LGSLOGEGFLTVHGDIVKRV 901

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10768511/runat_02122005.102334.18404/app.query.fasta_1.1095
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10768511 @CGN 1 1 5186 @runat_02122005.102334.18404 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :				GenEmbl.*			
				1: gb.ba.*			
				2: gb.in.*			
				3: gb.env.*			
				4: gb.om.*			
				5: gb.ov.*			
				6: gb.pat.*			
				7: gb.ph.*			
				8: gb.pr.*			
				9: gb.ro.*			
				10: gb.sts.*			
				11: gb.sv.*			
				12: gb.un.*			
				13: gb.vi.*			
				14: gb.htg.*			
				15: gb.pi.*			
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.				SUMMARIES			
Result				Query			
No. Score Match Length DB ID Description							
1 4609 100.0 4348 6 AR490068				AR490068 Sequence			
2 4609 100.0 4348 6 AX281543				AX281543 Sequence			
c 3 2016 43.7 110000 14 CR954214_1				Continuation (2 of			
				ALIGNMENTS			
RESULT 1				4348 bp DNA linear PAT 15-MAY-2004			
AR490068				Sequence 5 from patent US 6710229.			
LOCUS							
DEFINITION				AR490068			
ACCESSION				AR490068			
VERSION				AR490068.1 GI:47257191			
KEYWORDS				Unknown.			
SOURCE				Unknown.			
ORGANISM				Unclassified.			
REFERENCE				1 (bases 1 to 4348)			
AUTHORS				da Costa e Silva,O., Bohnert,H.J., van Thiel,N., Chen,R. and Sarria-Milan,R.			
TITLE				Cell cycle stress-related proteins and methods of use in plants			
JOURNAL				Patent: US 6710229-A 5 23-MAR-2004;			
				BASF Plant Science GmbH, Ludwigshafen;			
				WOX;			
FEATURES				Location/Qualifiers			
source				1..4348			
				/organism="unknown"			
				/mol_type="genomic DNA"			
ORIGIN							
Alignment Scores:				1.23e-162 Length: 4348			
Pred. No.:				4609.00 Matches: 901			
Score:				100.00% Conservativity: 0			
Percent Similarity:							

Continuation (202
AP004232 Oryza sat
AC116551 Dictyoste
AC007047 Arabidops
X58824 Schizosacch
AC149637 Medicago
U14731 Saccharomyc
249919 S.cerevisia
Z11255 S.cerevisia
D26089 Mus musculu
AK220442 Mus muscu
Continuation (3 of
CQ981174 Sequence
AX333086 Sequence
AX333087 Sequence
AX333494 Sequence
X74794 H.sapiens P
G28594 human STS S
BC074670 Xenopus t
BC031061 Homo sapi
BC013094 Mus muscu
BD159700 Primer fo
AX882395 Sequence
AK022899 Homo sapi
U46131 Xenopus lae
BC072870 Xenopus l
Continuation (14 o
BC083031 Xenopus l
U29178 Xenopus lae
Continuation (9 of
Continuation (16 o
Continuation (7 of
BC065958 Danio rer
AX281814 Sequence
U4049 Xenopus lae
Continuation (8 of
CQ17883 Sequence
BC056514 Danio rer
AX489180 Sequence
AC129365 Rattus no
AC119094 Rattus no
AK114961 Ciona int

us-10-768-511-8.rge

Thu Dec 8 10:36:56 2005

Best Local Similarity:	100.00%	Mismatches:	0	341	ProThrArgCysGluValProGluCysAlaAaArgAsnAlaMetSerLeuLeuHisAsn	360
Query Match:	100.00%	Indels:	0	2574	CCAACAAGGTGTGAAGAAGCCAGATGTGCAGCAGGATGCTATGCTCTCTATTATCAAT	2633
DB:	6	Gaps:	0	361	ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro	380
US-10-768-511-8 (1-901) x AR490068 (1-4348)						
QY	1	MetGluAsnAenAepAlaLeuAepIleGlyAlaValSerSerProTyrProSerGlnSer	20	2634	CGATGCCTTTTGCAAAATGAAGCAGATAGTGCCTCTTCAAGAACTCCAGATGCCATTCCT	2693
Db	1554	ATGGAAATATATGATGACATTGACATTGGAGCGGTGCTGCCCATATCTTCGCAATCT	1613	381	GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal	400
QY	21	GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAaSer	40	2694	GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACACACTATGTTGATGCTGTG	2753
Db	1614	GAAGGAGTGTCTACGCCATTCCGCCAAGTAACATCACCGAGCTTCGACNATGCGACCTCA	1673	401	LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly	420
QY	41	ProValAlaGlyArgCysAlaValArgGlnThrProThrSerAlaValArgArgGly	60	2754	AAGCCTGGAGATCGTATTTGAGGTAACAGGAGTTTCAAGGCCATGCGAGTTTCGATGGT	2813
Db	1674	CCCGTGGCGCGGAGGCGGTACCGCAGACCCCTACATCTCGAGTTTCGAAGGAGAGGG	1733	421	ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys	440
QY	61	ArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerVal	80	2814	CCGAATCAACGAACATTTACGAGCATTTGTATAAGACCTACATCGATTCGTCACGTCAAG	2873
Db	1734	AGAGAAACGGATTCGCTCGTGTAGGAGAGTCGATCTCGCAGTTTAGGCAATTTCTGTT	1793	441	LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp	460
QY	81	TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro	100	2874	AAGTCTGACAGGGTTCGACTGCAAACTGAAGATCTCTATGGAGATGGATAAGGAGATGAT	2933
Db	1794	TATAGTTCCCTTACGATCGGGGACTCCTGGAACCTCCTGGAACTCCAGTGCTACTCG	1853	461	MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln	480
QY	101	ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro	120	2934	ATGTATGCTGGGTATCATGAAGTGATCTTTCAGAAAGCTGCTTAATGAAGCAAGATTCAA	2993
Db	1854	GTTTACGCTACCCACGTCGGTACACTATGGGTACCCCATCGTTCATCGTGCGACGCCA	1913	481	LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla	500
QY	121	GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysPheLeuHisArgArgArg	140	2994	AAACTTAAAGAGCTGTCCAGCTCCCGGCAATTTATATAGATAGACTTTTCAAGGTCGCT	3053
Db	1914	CAGTACAAACGCGCAGTGAGCTTGGTTCCAGGGGAAGCTCTATACGAGACGTCCA	1973	501	ProSerIleTyrGluLeuGluAspIleLysGlyLeuLeuCysGlnLeuPheGlyGly	520
QY	141	SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg	160	3054	CCDAGCAATTTGGAGCTTGAAGATATTAAGAGGGTCTCTTTTCCAGCTCTTTGGTGGG	3113
Db	1974	TCTCAATCCAGAGAACCGGGCATCGATCTCTTCAAGGGGAACCTTAGTGTGATGGCGT	2033	521	LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuVal	540
QY	161	ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGlyTyrAlaTyrValTyrGly	180	3114	AAGCTAAGAAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGT	3173
Db	2034	CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGNATATGCTTATGTTGGGG	2093	541	GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg	560
QY	181	ThrAsnValAsnIleProAspValLeuArgAlaIleArgPheLeuHisAsnTyrArg	200	3174	GGGGACCTCGTACAGTAATCTCAGCTGCTTCAGTATGTGCACAAGATAGTCTCTCGT	3233
Db	2094	ACGAATGTTACATTTCCAGATGTGCTTAGGGCGATTCTGCGATTTCTCCACATATATCGT	2153	561	GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys	580
QY	201	SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleGluGluThrValGluArg	220	3234	GGAAATCTACACTAGTGGCGAGGAAGTTCCGGCGTTCGGCTGACACAGCTATGTAACGAAG	3293
Db	2154	TCGAGTGTCTCATGATCTTAATTCGAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT	2213	581	AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly	600
QY	221	GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr	240	3294	GATCCAGAAACTCCGAGAGACCGTATTTGGAGAGCGGAGCTTTGGTCTTTAGTATCGTGGG	3353
Db	2214	GAGGAGGATACCTAAATATCCAGATGTCAGCATTTATGACCATGATCTCTATATAC	2273	601	IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaAArgSerMetLeuHisGlu	620
QY	241	AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu	260	3354	ATATGCTGTATTCATGAGTTCGACAAAATGTCTGATAATGCCCGGAGCATGCTTCATGAG	3413
Db	2274	GCAAAAAATTTGTTCCATACCCACTCGACATCATCCCTCTGTGGACACTGAGTGTGAGAA	2333	621	ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla	640
QY	261	ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu	280	3414	GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTTGCTCGTGAAGCT	3473
Db	2334	GTTTGCTACCTTTTACTACCAAGTTTTGAAGACATATTGAGGCCAGACCTTTCAATCTC	2393	641	ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer	660
QY	281	LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal	300	3474	CGGACGCTGTCTTCATGTGCAATTCCTAGTGGGTCCCGATACAAATGCGCGCTTTCT	3533
Db	2394	AAAGCATCGGTGCACATGCGTGAACTCAACCTTCAGATATAGACAAATTTGTTCTGTT	2453	661	ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu	680
QY	301	LysGlyMetValIleArgCysSerSerIleIleProGluLysGlyAlaPhePheLys	320	3534	GTGATGTGATTAACATCCAGCTTCTCCAACTCTACTTCTAGATTTGATTTAATTTACTTA	3593
Db	2454	AAAGGAATGGTATATCCCGGTGAGTTCATCATCATCTGAAATTAAGGGGGCTTCTTCAA	2513	681	MetLeuAspLysProAspGluGlnAsnAspArgLeuAlaArgHisLeuValAlaLeu	700
QY	321	CysLeuValCysGlyHisSerProProLeuValThrValLysGlyValGluGlu	340	3594	ATGCTCGACAAACGAGACGACAAACGATCGTCTTCGCCAGGCATCTCGTGGCTTTA	3653
Db	2514	TGTTTAGTGTGTGCTCACTCCCTCCCTAGTTACAGTTGTTTAAAGGGCGGGTTGAGGAG	2573			


```
QY 72 ArgSerArgSerLeuGlyAasnSer----- 79
Db 73092 CTCGATCGATCGATGGCGAAGCGACAGCTGTGGATGAAACGTCGCCCGCGACGGCGA 73033
QY 80 -----ValTyrSerSerProTyrAspAla 87
Db 73032 GGGACGAGTGCAGGGGGCTGGGGCGATGTGGGGTTGATGTCGCCCGCGCGCGCGC 72973
QY 88 GlyThrProGlyThrProGly-----ThrProValAlaThrProValTyrAla 103
Db 72972 GAGCGAGCTCGGGGGTGCAGGAGGATGTGCGCTGTGGATCGGATTTGTTTCC 72913
QY 104 ThrPro-----ValGlyThrProMet-----Gly 111
Db 72912 CGCGCGGAGACGCGCGACGCGCGGCTCCACGCGATGGTGGTGGAGGCTTTGGA 72853
QY 112 ThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGln 131
Db 72852 TCACCGCGTTC-----GGGACGCGAGACGACGCGCGAGGGGATGTT---CGACGA 72802
QY 132 GlyLysProLeuHisArgArgArgSerGlnSerArgGluProGlyHisArgSerPro 151
Db 72801 GGGGTGTTGAGTGCAGCGCGGACGAGTGGATGAGTGGAGGACGATGGATTCGCGCGCGC 72742
QY 152 SerArgGluProSerAlaAspGly----- 159
Db 72741 GGGAGCGACGCGCGGCGAGCGCGGCTTGGACGCGACTCTCTGGAACGCGAGTTGACGGCG 72682
QY 160 -----ArgProSerGluSerAlaGluProAspAspThrLeu 171
Db 72681 ACGCAGGATGTTGATTTCTTGTGATGTACCGCGCGAGAGCGCG---GGTGATCACTTGTATG 72625
QY 172 GlyGlyGlu-----TyrAlaTyrValTyrGlyThrAsnVal 183
Db 72624 CTGGGCAATGATCGCAGGACGTGGCGATGCGACGATATATCTGGGGTACCAAGGTG 72565
QY 184 AsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg----- 200
Db 72564 AACGTGTACAGCTGCAGACGCGCTTCGCGGATTTTGGMAAACTTTGANGTGGTATGTC 72505
QY 200 ----- 200
Db 72504 CTTCCGTGCGAGTTTCAGTGGCTGAGGGCTACGGTGGAGCTGATTTGTGCTCGCGTT 72445
QY 201 -----SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIle 213
Db 72444 GATTCGTTCCGCCACCGACTCGTAACATCCACGACTG----- 72406
QY 214 IleGluGluThrValGluArgGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
Db 72405 ATGCAGATCTCTGGAGAGAGAGCAATCGGATCTGGATCTGGATCTGATGTCAGCAGTCGAC 72346
QY 234 AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu 253
Db 72345 AAATACGATCCGTTCTTGACAGTCTTTCGACGAGTACCAGGAGATCAATTCGCTC 72286
QY 254 LeuAspThrGluCys-----GlnGluValAlaThrSerLeuLeuProThr 268
Db 72285 TTCGAGCTGTTGCCAAATGAATATCTTACGCCAAATGTCTGTGCCAGAGTGTGCGCGGT 72226
QY 269 PheGlu-----LysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHis 285
Db 72225 GATGAGCCGATGATCACCGATTTGTAGTCGCCCGCTTCAACATGATGGAGGCGAAACCC 72166
QY 286 MetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIle 305
Db 72165 ATCGCGACTTGAATCCATCGGACATCGACAAGATGTGTGCGTGGTGTATGTTGACG 72106
QY 306 ArgCysSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly 325
Db 72105 CGATGCACACGATCATTTCTGACCTAAACTCGCGTATTTCAAGTGTCTGATGTGCGCG 72046
```

```
326 HisSerProProLeuValThrValValLysGlyArgValGluGluPro----- 341
72045 TTCGCCCAAGACATATTCAGATCCGCGGTCGCTCAACAGACCTCCGCTCAAGTGC 71986
342 ThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArg 361
71985 ACAGAGTGGGCAAGCCA-----GNAACGATGACCTTGATTCACATCAA 71941
362 CysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGlu 381
71940 TGCCTGTCGCAAAACAAGCAACGGTGAAGATGCAGGAGACACCGATCGATCCGAG 71881
382 GlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLys 401
71880 GGTGAGACTCCGACACTGTATCCATGCTGCTTCGATGATCTCGTTGATCAGGCGAAG 71821
402 ProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyPro 421
71820 CCAGGTGACCGAGTTGAAGTTACTCGAGTGTATCGTGGTTCGAATTCGTCTGAGCTCT 71761
422 AsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLys 441
71760 ACGAGCGCACGCTCAAGAGTGTGTACAAGACGTACTAGACGTTTTTTCATATAAGAAAG 71701
442 SerAspArgGlyArgLeuGln---ThrGluAspProMetGluMetAspLysGluAsnAsp 460
71700 GAGCTCGGGCGGAGGATCGAAACACGCGCGGACCC-----GAGGAT 71659
461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
71658 GAAGAGGCTGCAAGAAATCAGCTGCTTCAACGAATCATCTGGACCGGTGAAGAATCAA 71599
481 -----LysLeuLysGluLeuSer 486
71598 GGTCTCGGACCAATGAGTTCTACTCCGAGCGCATCGCGGACAGATAGAGAGCTCGT 71539
487 LysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrGluLeu 506
71538 CGTTCGCGGACATTTATGACCGGCTGTGGCGTCTGTGGCCCGCTCAATTTGGGAGTTA 71479
507 GluAspIleLysGlyGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysIlePro 526
71478 GAAGACGTTTAAAAAGGACTGTTGTGCTAGCTTTTGGCGCCACGAAAGAGCTTCAGC 71419
527 SerGlyAlaSer-----PheArgGlyAspIleAsnValLeuLeuValGlyAspProGly 544
71418 GACAAGCAGCGCAACAGAGTCGCGGCGACATCAATTTTACTCGTCGTCGATCCGCG 71359
545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
71358 GTCGCCAAGTCTCAGCTATTGACGTACGTTTCTATCGTATCGCGCGCTGATGTACACC 71299
565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
71298 TCAGGTAGAGGTCGTCGCTGCTGCTTACAGCGTACGTCTCAGTCGCGATCCGAGTCA 71239
585 ArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
71238 AAAGCATGCTGCTCGAGTCAGAGCGCTCGTCTTCTGATCGCGGCTTTGTCGATTT 71179
605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
71178 GATGAGTTCGACAAAGATGCTGACAGTGTCTGATCAATGTTGACGAAGTGTAGGACAG 71119
625 GlnThrValSerValAlaLysGlyIleIleAlaSerLeuAsnAlaArgThrSerVal 644
71118 CAGACGTTTCCATCCGAAGGCTGGAATCATTTGCGGTGTAAATCAGCGACTTCGGTC 71059
645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
71058 CTTGCTCAGCGAACCCAGTCGGAAGTCTTACACCCCGAACATGTCATGTTGGAGAAC 70999
665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
```

```
Db 70998 ATTCACTCCGCCGACACTGTTGCCGTTTGGTTGCTACTATGCTTGCACCGT 70939
QY 695 ProAspGluGlnAsnAspArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
Db 70938 CCTAATCCGAGACGATCGCGTCTGGCGAGACACTTAGTCAGTTTGCACACAAAT 70879
QY 705 TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuAlaTyrIleThrTyr 724
Db 70878 CCCCGCAGAAAGCGGCGCTGATTTCTGCCGACTTCTCACCAGTACGTCCTGAT 70819
QY 725 AlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGly 744
Db 70818 GTCGCGCAAGTACACCGGTGTCAGCGACGAGCTCGAGAGAGCTCGTGAGGGA 70759
QY 745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
Db 70758 TATGTTGAGATCGGACGATG-----GGTGGTAGTCGAAAGTTATTACTGCA 70711
QY 765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
Db 70710 ACGCCGAGACAATTGAGTCTCTCATTCGTCTCAGAGTCTCTAGCGCGCATGCGTTTA 70651
QY 785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
Db 70650 AGTCGGTGTGGATCGTGACACGCTAAGGAGGCGTGCAGATTATGCGGGTGGCATG 70591
QY 805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrGly 824
Db 70590 CAACAATCGCGCTGATCCTCGCACTGGGACGATTGATGATGATAAGATATTGACCGGA 70531
QY 825 ValSerAlaSerGluArgIleArgAlaAsnLeuLeuAlaLeuArgGluLeuIle 844
Db 70530 CATAGCGCAGACCGCTCAGACCGCAGCGGTGCGAGAGCCATCAGGCGGTGT--- 70474
QY 845 AlaAspLysIleSerProGlySerSerGlyLeuLysThrSerGlnLeuLeuGluAsp 864
Db 70473 -----CTTAATACTCTGGAGCGGTAGACTCCGACTGGCGGAGTCGCGAAGCG 70423
QY 865 IleArgSerGln---SerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGly 883
Db 70422 CTCGCGGAGAGGACAAACAGTGGCGTGGCATCCAGGAGGTGCGGACCGCGCAATG 70363
QY 884 SerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
Db 70362 CTTCCTGTAGAGAGGAAAGTGTACCATTCGAGGTGACCTCGTC 70318

RESULT 4
AP008207_201
WPCOMMENT
Sequence split into 433 fragments LOCUS AP008207 Accession AP008207
Fragment Name Begin End
AP008207_000 1 110000
AP008207_001 100001 210000
AP008207_002 200001 310000
AP008207_003 300001 410000
AP008207_004 400001 510000
AP008207_005 500001 610000
AP008207_006 600001 710000
AP008207_007 700001 810000
AP008207_008 800001 910000
AP008207_009 900001 1010000
AP008207_010 1000001 1110000
AP008207_011 1100001 1210000
AP008207_012 1200001 1310000
AP008207_013 1300001 1410000
AP008207_014 1400001 1510000
AP008207_015 1500001 1610000
AP008207_016 1600001 1710000
AP008207_017 1700001 1810000
AP008207_018 1800001 1910000
AP008207_019 1900001 2010000
AP008207_020 2000001 2110000
AP008207_021 2100001
AP008207_022 2200001
AP008207_023 2300001
AP008207_024 2400001
AP008207_025 2500001
AP008207_026 2600001
AP008207_027 2700001
AP008207_028 2800001
AP008207_029 2900001
AP008207_030 3000001
AP008207_031 3100001
AP008207_032 3200001
AP008207_033 3300001
AP008207_034 3400001
AP008207_035 3500001
AP008207_036 3600001
AP008207_037 3700001
AP008207_038 3800001
AP008207_039 3900001
AP008207_040 4000001
AP008207_041 4100001
AP008207_042 4200001
AP008207_043 4300001
AP008207_044 4400001
AP008207_045 4500001
AP008207_046 4600001
AP008207_047 4700001
AP008207_048 4800001
AP008207_049 4900001
AP008207_050 5000001
AP008207_051 5100001
AP008207_052 5200001
AP008207_053 5300001
AP008207_054 5400001
AP008207_055 5500001
AP008207_056 5600001
AP008207_057 5700001
AP008207_058 5800001
AP008207_059 5900001
AP008207_060 6000001
AP008207_061 6100001
AP008207_062 6200001
AP008207_063 6300001
AP008207_064 6400001
AP008207_065 6500001
AP008207_066 6600001
AP008207_067 6700001
AP008207_068 6800001
AP008207_069 6900001
AP008207_070 7000001
AP008207_071 7100001
AP008207_072 7200001
AP008207_073 7300001
AP008207_074 7400001
AP008207_075 7500001
AP008207_076 7600001
AP008207_077 7700001
AP008207_078 7800001
AP008207_079 7900001
AP008207_080 8000001
AP008207_081 8100001
AP008207_082 8200001
AP008207_083 8300001
AP008207_084 8400001
AP008207_085 8500001
AP008207_086 8600001
AP008207_087 8700001
AP008207_088 8800001
AP008207_089 8900001
AP008207_090 9000001
AP008207_091 9100001
AP008207_092 9200001
AP008207_093 9300001
```

AP008207_094	9400001	9510000	16700001	16810000
AP008207_095	9500001	9610000	16800001	16910000
AP008207_096	9600001	9710000	16900001	17010000
AP008207_097	9700001	9810000	17000001	17110000
AP008207_098	9800001	9910000	17100001	17210000
AP008207_099	9900001	10010000	17200001	17310000
AP008207_100	1000001	10110000	17300001	17410000
AP008207_101	1010001	10210000	17400001	17510000
AP008207_102	1020001	10310000	17500001	17610000
AP008207_103	1030001	10410000	17600001	17710000
AP008207_104	1040001	10510000	17700001	17810000
AP008207_105	1050001	10610000	17800001	17910000
AP008207_106	1060001	10710000	17900001	18010000
AP008207_107	1070001	10810000	18000001	18110000
AP008207_108	1080001	10910000	18100001	18210000
AP008207_109	1090001	11010000	18200001	18310000
AP008207_110	1100001	11110000	18300001	18410000
AP008207_111	1110001	11210000	18400001	18510000
AP008207_112	1120001	11310000	18500001	18610000
AP008207_113	1130001	11410000	18600001	18710000
AP008207_114	1140001	11510000	18700001	18810000
AP008207_115	1150001	11610000	18800001	18910000
AP008207_116	1160001	11710000	18900001	19010000
AP008207_117	1170001	11810000	19000001	19110000
AP008207_118	1180001	11910000	19100001	19210000
AP008207_119	1190001	12010000	19200001	19310000
AP008207_120	1200001	12110000	19300001	19410000
AP008207_121	1210001	12210000	19400001	19510000
AP008207_122	1220001	12310000	19500001	19610000
AP008207_123	1230001	12410000	19600001	19710000
AP008207_124	1240001	12510000		
AP008207_125	1250001	12610000		
AP008207_126	1260001	12710000		
AP008207_127	1270001	12810000		
AP008207_128	1280001	12910000		
AP008207_129	1290001	13010000		
AP008207_130	1300001	13110000		
AP008207_131	1310001	13210000		
AP008207_132	1320001	13310000		
AP008207_133	1330001	13410000		
AP008207_134	1340001	13510000		
AP008207_135	1350001	13610000		
AP008207_136	1360001	13710000		
AP008207_137	1370001	13810000		
AP008207_138	1380001	13910000		
AP008207_139	1390001	14010000		
AP008207_140	1400001	14110000		
AP008207_141	1410001	14210000		
AP008207_142	1420001	14310000		
AP008207_143	1430001	14410000		
AP008207_144	1440001	14510000		
AP008207_145	1450001	14610000		
AP008207_146	1460001	14710000		
AP008207_147	1470001	14810000		
AP008207_148	1480001	14910000		
AP008207_149	1490001	15010000		
AP008207_150	1500001	15110000		
AP008207_151	1510001	15210000		
AP008207_152	1520001	15310000		
AP008207_153	1530001	15410000		
AP008207_154	1540001	15510000		
AP008207_155	1550001	15610000		
AP008207_156	1560001	15710000		
AP008207_157	1570001	15810000		
AP008207_158	1580001	15910000		
AP008207_159	1590001	16010000		
AP008207_160	1600001	16110000		
AP008207_161	1610001	16210000		
AP008207_162	1620001	16310000		
AP008207_163	1630001	16410000		
AP008207_164	1640001	16510000		
AP008207_165	1650001	16610000		
AP008207_166	1660001	16710000		

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

5.26e-62

1937.50

41.94%

33.52%

42.04%

15

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

110000

530

133

173

751

31

US-10-768-511-8 (1-901) x AP008207_201 (1-110000)

QY

11

AlaValSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnVal

30

Db

88290

GCCGTGTCGTCCGCG

-----GATGTGCGACCGTCGAGCCCTCTCCGCGCGAC

88337

QY

31

ThrSerProSerPheAsnAlaValSerProValAlaGlyArgAlaValArgGln

50

Db

88338

-----AACTCTCCCGCCTCAGTCGGCGCGCGCGC

-----88370

QY

51

ThrProThrSerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArg

70

Db

88371

-----GGGGGGCGCGCGCGGT

88388

QY

71

SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro

90

Db

88389

GCTTCC-----GCCAGCCCGTACCGTCG

---TCCCG 88418

QY

91

GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet

110

Db

88419

TCCCTCGCGCGG-----TTCGAGACCGCGCGCACCGCGCGCGC

88457

QY

111

GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySer

130

Db

88458

CCACGCCCTCC-----GGCGCGCGCGCGCGCAGCAGCGGCA-GAATGGACGG

88510

QY

131

GlnGlyLysPro-----LeuHisArgArgArgSerGlnSer

143

Db

88511

GGGGCGGTTCGCCCGCGACCCCGTCCACCCCATGTCTCCACGACGACGTC-CGCTCTCCT

88569

QY

144

ArgGluProGlyHisArgSerProSerArgGluProSerAla--AspGlyArgProSerG

163

Db

88570

CGGAGCGCGGACGAGGACACCCCGACCGACCGCGCGCGCGCGCGCGCGCG

88629

QY 163 luSerAlaGluProAspAspThrLeuGlyGlyLeuTyAlaTyAlaTyValTrpGlyThrAsnV 183
Db 88630 CCGAGCCACGGCG-----GTGTTCTGTGGGGCACCACA 88665
QY 183 alAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyArgSerSera 203
Db 88666 TCAGCGTGCAGGACGTCAACGCGCCATCTCCGGTCTCTCGCCACATTCGCG---GACC 88722
QY 203 laHisAspLeu-----AsnSerLysTyTrileGlnIleIleG 215
Db 88723 CCGCGAGCGCGCGCGCTGCACCGGTCAATGACAGCGGCAAGTACATGCGCGCATCC 88782
QY 215 luGluThrValGluArgGlu---GluAspThrLeuAsnIleAspMetSerAspIleTyra 234
Db 88783 ACCGATCTCTGAGTCTGAGCGCGGGAGTCTCGAGTCAACGCCCCACGACGTGTCG 88842
QY 234 sPhisAspProAspLeuTyAlaLysIleValArgTyProLeuAspIleIleProLeuL 254
Db 88843 ACCAGCCCGGACCTCTACGCCAAGATGGTCAAGTACCGGTCTGAGGTCTCGCCATCT 88902
QY 254 euAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleG 274
Db 88903 TTGATCGTGTCTATGAGCTCTGCGCGCATCGAGCCCTCTTCGAGAAGCACATCC 88962
QY 274 luAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer---- 292
Db 88963 AGACCGAGATCTACACCTCAAGTCTCGTTCGTTGCTTGAAGATCTCAACCCCTTC-TGGT 89021
QY 292 ----- 292
Db 89022 GATTGGCAGCTCTGCTCTCTTTAGTACTTGTGTCGGCAATCTTTTATATGATTC 89081
QY 293 -----AspIleAspLysLeuValSerValLysGlyMetValIleA 306
Db 89082 TGATTTTGGTTCGTTGTCAGATCGAGAGATGGTTCATCAAGGTATGATAATTC 89141
QY 306 tqCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly- 325
Db 89142 GGTGAGCTCGGTGATTCAGAGCTCAAGAGGCTGTGTCGCTGCTTGGTTCGGGT 89201
QY 326 --HisSerProProLeuVal----- 331
Db 89202 TCTACTCTGAGCTGTATGGTGTGATCGAGGTATATATATATATATATATATATATAT 89261
QY 331 ----- 331
Db 89262 CTTTTTCACACTGCGTAATTCATATGCTCTTGTGTTAAGTGAATTCAGGTGATCAATTA 89321
QY 331 ----- 331
Db 89322 ATCAGATTGGAGCGCTTAATTTGCTCATAGAGTTACATATTGACAACCTCAATTGTAA 89381
QY 331 ----- 331
Db 89382 CTTTCATTGTACAAATCATCTTATCATGAAAGATGATAAATTCAGCTTCATTGACAT 89441
QY 331 ----- 331
Db 89442 GTTATATGTCACTACTATTTATCATTTATGCAATTTCTTTTATTTTGTGCAAAAACTTT 89501
QY 331 ----- 331
Db 89502 TCATTTTGTGGGATACACAACATTTAGGAATTTTATTTATCATACCAATTCGTGAGGA 89561
QY 331 ----- 331
Db 89562 ACTGATAAATAGTATATAATTGATAGCGCAAAATCAACATCCATCTACTGCTATTGTCCA 89621
QY 331 ----- 331
Db 89622 ATCCAAGACTCGGACTAGTCTGTCTGTATATAAGTAACACATTTGCACTGTGGTAGTAAT 89681
QY 331 ----- 331

Db 89682 TGTAAAGTAAACACATGCTTATGTTCTTAACTTACATATATTTCTGGTGGTTAATGTTTGA 89741
QY 331 ----- 331
Db 89742 AGAGATTGGTTCATGCAGATCCTATGAAAAACACATTAGTGGCCTAGTGGGAGTAATGACTA 89801
QY 332 -----ThrVal-ValLysGlyArg 337
Db 89802 AAATCCCATCAGTTTCTGTTGTTTATCATGTAATAATTCACTCTATTTCTTAGGAGG 89861
QY 338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db 89862 GTTACTGAGCGCCACATCTGTCAAGAAACAAATGTAAGCTACAACTCTATGACTCTT 89921
QY 358 IleHisAsnArgCysThr----- 363
Db 89922 GTGCATAACCGATGCAG-GTAGCTATATCTCTACTTGTAAACCCCTTTTCTTTATAGGATT 89980
QY 364 -----PheAla 365
Db 89981 TTTGACAGATATGATGCTCTGACCTTTGGATGGAACATTTTATTGTAACTAGGTTTGA 90040
QY 366 AsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPro 385
Db 90041 GATTAAGCAGATCATTAAGTTGCAGAAACACACAGATGATGATCCAGAGGTGGCACTCA 90100
QY 386 HisThrValSerMetCysLeuTyArgAsnThrMetValAspAlaValLysProGlyAspArg 405
Db 90101 CATACTGTGTAGTCTTGATGCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGG 90160
QY 406 IleGluVal----- 408
Db 90161 GTTGAGGTTAGCAGTGAATCTTTGTTGGTCATTAATAACAAGAAATAATTTGTTGTG 90220
QY 409 -----ThrGlyVal 411
Db 90221 GTTCTTATTCATCGCTAAAAATCTGATTTTTTCTTTGATTTTATAGATAACTGGATA 90280
QY 412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg----- 427
Db 90281 TACAGAGCCATGATGATTAATGAGTTGGCCCAACTCAGAGACAGTGAAGTCGATTTTCAAG 90340
QY 427 ----- 427
Db 90341 GTAGAGGATGTCAATGTTATCTTACTTGTCTATCTTAAATTCCTTTCTGCACAGAAAA 90400
QY 428 -----AlaLeuTyLysThrTyTrile 434
Db 90401 AAGTTTTAAATATCTACTCAAGCATCGCTAACAGTTTTCTTTAAT-CAGACGTACATT 90459
QY 435 AspCysValHisValLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db 90460 GATTGCTTTCATAAAGAAGACAGACAAGTCTAGACTTAGACTTTCATGTTGAGGACTCCATGGA 90519
QY 455 MetAspLysGluAsn-----AspMetTyAla 463
Db 90520 ACTGATAACCCCAATGCTAACAGACAACCTGAAGATGATTTTCTCAGAGATAAGGTATCC 90579
QY 464 GlyTyHisGluSerAspThrSerGluAla-----Ala 474
Db 90580 CTCATCACTTCTCTCTCGGATA-GCACTAAGTGAATGTGTTTAACTGCTCCTT 90638
QY 475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyArgAspArg 494
Db 90639 TCTGTATTGACAGTTGAGAAATTAAGAGATTGCAAGATTGTCAGAGATATATATATACAGA 90698
QY 495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysGly----- 512
Db 90699 TTAAGTAGTCTTGGCTCCAAACATATGGGAGCTGGACGACGTTTAAAGGGGCTCCTT 90758
QY 512 ----- 512

```
Db 90759 TGCCAGGTAAGATCATCTATCTGTTTGAATGTTTTTGCACATAATAGCCTAGTATACAAG 90818
Qy 513 -----LeuLeuCys-GlnLeuPheGlyGlyLysAlaLysIleProSerGlyAlaSe 530
Db 90819 CTCCTACTGTTGTGTCAGCTTTTGTGGAAATGCTTTGAGGCTCTCTCTGGAGCTAG 90878
Qy 530 rPheArgGlyAspIleAenValLeuLeuValGlyAspProGlyThrSerLysSerGlnIle 550
Db 90879 TTTCCGAGCGACATCAATATATTTGCTTGTGATCTGCTGGAACAAGTAATATCCAGCT 90938
Qy 550 uLeuGlnTyrValHisIleLeuAlaProArgGlyIleTyrThrSerGlyValArgGlySerSe 570
Db 90939 TCTCAATACATGCAACAACCTCTCTCGTGGCAATTTATACAAGTGGCAGAGGAAGTTC 90998
Qy 570 rAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr- 587
Db 90999 AGCTGTGTCCTTACTGTCATAGTTTACCAGGATCTGGAACCTGGTGAACCTGTAAGTCT 91058
Qy 587 ----- 587
Db 91059 ACATGAATCCTTTATTTTGCACACTTAATTTAACTCGTTTAAATGTAATCTGATTGATGA 91118
Qy 588 -----ValLeuGluSerGlyAlaLeuValLeuSerAspAr 599
Db 91119 CTTCAACCATGATCATCTCAATTTAGGTTCTTGAGAGTGGAGCGCTTTGTTTGGTGACAA 91178
Qy 599 gGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaAtrSerMetLeuHi 619
Db 91179 AGGTGTGTTGTTATGTAATTTGTAAGATGTCGTATATGTCGGAAGCATGTTACA 91238
Qy 619 sGlu----- 620
Db 91239 TGAGGTCTCTTGTATGCTGCGCCAGAAATTTGTATCACACAGTACTATTATTTAGTCAATG 91298
Qy 621 -----ValMetGluGlnGlnThrValSerValAlaAl 631
Db 91299 CTAACATGTCCTTTTTCAAAAATATTGACAGGTGATGGAAACAACAGACTGTCTCCATTGCCA 91358
Qy 631 yGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProS 651
Db 91359 AGGCTGGAATAATGTCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCMAATCTTA 91418
Qy 651 exGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThrL 671
Db 91419 CTGMAATCAGTTATATATCCAGGCTCTCTGTGATTGACAAATATCCATCTCTCTCCACAC 91478
Qy 671 euLeuSer----- 673
Db 91479 TGCTTTCTAGGTGAGTGCATATGCAATGTGAATGAGCAATTTCTTTCAGGACAAACAGATCA 91538
Qy 674 -----ArgPheAspLeu 677
Db 91539 ATAATAGAGTGTTTAACTCTTTGATGTTTTTATTTTTTCAAAATCAGTTTGTACCTC 91598
Qy 678 IleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeu 697
Db 91599 ATTATCTGATATTGACAAGGAGATGACAAATGATAGACGCTGCTGGCTAGCATATT 91658
Qy 698 ValAlaLeuHisTyrGluLeuTyrGluVal-SerLysGln- 710
Db 91659 GTTTCGTGTGATTTTGAGAAATCCAAACGTAAACACAAACAACACTCCTTTCTTCTCTGAAGTT 91718
Qy 710 ----- 710
Db 91719 TGTACTTTTTTACTTATTTGATATAAAGTCTAAACCGTCTTGCAAAATCAGAACACTGTTG 91778
Qy 710 ----- 710
Db 91779 TTAGCTTTAGCTATTCTTATTAATTTGTCATCTCTTCTGCTATCTAGATAGGAGCT 91838
Qy 711 -AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHi 730
Db 91839 CGAGGTCTTGGATTTGGCAACACTTGTAGCTTACATAAGTTATGCAAGGAAGCATATACA 91898
```

```
Qy 730 sProThrLeuSerAspGluAlaAlaGluAspLeuIleAenGlyTyrValGluMetArgGl 750
Db 91899 ACCACAGTTATCTGATGAAGCTGCAGAAAGAAATTTGACTCGCGCTATGTTAGATGAGCAA 91958
Qy 750 nLysGlyAenPheProGlySerSerLysLys- 760
Db 91959 AAGAGAAACAGCCCTGTTAGCAGAAAGAGGTAGTGTAGGCTTTTTTAATATGATATTTT 92018
Qy 761 -----ValIleThrAl 764
Db 92019 TAGCCATTAAATGGTACAGTGACTGATGCTTTTTTGTCTTATTTGATAGGTCTATAACTGC 92078
Qy 764 aThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPh 784
Db 92079 GACAGCTCGACAAATTTGAGAGCTTGAATTCGGCTCAGTGAAGCACTGGCCCGAATCGATT 92138
Qy 784 sSerGluVal----- 787
Db 92139 CTCTGAAATGGTAATGATCATTCCTTATTTCTTAGTTGTCTTGGCTTTCACAGCATTAATCT 92198
Qy 788 -----ValGluLysValAspAl 793
Db 92199 GGTACTTTTGGTACTCATGATTTCTTCTCACTTGTCTACAGGTTGAAGTACAAGATCT 92258
Qy 793 aAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaTh 813
Db 92259 TGTAGAGGCTTTCAGGCTTCTCGAAGTTGCCATGTCAGCAATCGCAACTGATCATGCCAC 92318
Qy 813 r----- 813
Db 92319 TGGTCAGACAATTTGCTTAGCGCCATTTCTTATGCAATTTTTTCTTGTAAACATTAACAT 92378
Qy 814 -----GlyThrIleAspMetAspLeuIleTh 822
Db 92379 AAAGTCTGACATTTTATCAATCAACCACTAAAGSTACAATCGATATGATCTTATCAT 92438
Qy 822 rThrGlyValSerAlaSerGluArgIleArgAlaAenLeuLeuAlaAlaLeuArgGl 842
Db 92439 GACTGGAATATCTCGAGCGAAAGGAGGCGGAGCAATCTCTGTTGCAGCAACCCGCAA 92498
Qy 842 uLeuIleAlaAspLysIleSerProGly----- 851
Db 92499 CCTTGTGATGGAGAAATATGCACTTTGGAGGGCCCTCAGTGCAGATGATTGAGTAAGAA 92558
Qy 852 -----SerSerSerGlyLeuLys----- 857
Db 92559 CCTTCAATTAGTGTCTTGTCTCATCATTTGTTGCCAAACCACTGGACTAAGTTATTTC 92618
Qy 858 -----ThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSe 874
Db 92619 TTCTCTGATTGATTTCGACGTTCTGGAAGAAATTAGAAGCAGACAGCTCTATGGAAGTTCA 92678
Qy 874 rLeuGlnAspIle----- 878
Db 92679 TCTGTCATGATGTAAGTTGGAGATCTAGTCTGCATTTCTTCATTTGTCATGACCTGATACCTGGG 92738
Qy 879 -----LysAsnAlaLeuGlySerLe 885
Db 92739 AGGCTCTAACCCATTCCTGAAATTTTCCATCTTCTCAGCTTTCGGGTGCTCTTGGCACTCT 92798
Qy 885 uGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
Db 92799 GATGACAGAGGGTGGCGGTAGTCATCCATGAGACAGCGTCAGAGGGGTC 92847
```

RESULT 5
AP004232
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
BAC clone: OSJNBa0051H17.
ACCESSION
AP004232 BAO00010
VERSION
AP004232.4 GI:20161778
KEYWORDS

SOURCE ORGANISM	Oryza sativa (japonica cultivar-group)
REFERENCE AUTHORS	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE	1
AUTHORS	Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Nimura, Y., Cheng, Z., Nagamura, Y., Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikono, M., Ito, S., Ito, T., Ito, Y., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katsagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohka, I., Ono, N., Sai, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yanagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y., Yano, M., Jiang, J. and Gojobori, T.
TITLE	The genome sequence and structure of rice chromosome 1
JOURNAL	Nature 420 (6913), 312-316 (2002)
REFERENCE	12447438
AUTHORS	2 (bases 1 to 187471)
TITLE	Sasaki, T., Matsumoto, T. and Yamamoto, K.
JOURNAL	Direct Submission Submitted (03-OCT-2001) Takuji Sasaki, National Institute of Agricultural Sciences, 305-8602, Japan 2-1-2, Tsukuba, Ibaraki, 305-8602, Japan (E-mail: tsasaki@ias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)
COMMENT	On Apr 16, 2002 this sequence version replaced gi:19110523. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/tdb/glimmer/glmr_form.html), RiceHMM (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OSJNBA0051H17 clone has an overlap with OSJNBA006517 (DBJ: AP004233) clone at 5' end and with OSJNBA002614 (DBJ: AP004231) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html. Location/Qualifiers 1..187471 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare"
db_xref="taxon:39947"	
/chromosome="1"	
/clones="OSJNBA0051H17"	
3101..4772	
/gene="OSJNBA0051H17.1"	
join(<3101..3168,3216..3273,3487..3700,4630..>4772)	
/gene="OSJNBA0051H17.1"	
/note="start and end point are not identified"	
join(3101..3168,3216..3273,3487..3700,4630..4772)	
/gene="OSJNBA0051H17.1"	
/note="predicted by GENSCAN etc."	
/codon_start=1	
/product="hypothetical protein"	
/protein_id="BAB90694.1"	
/db_xref="GI:20161779"	
translation="MWRLSAAGDVLVKIQEQASRPGVGQCPTPPVQASRRVVV IRREQVALINPTFYFADTHGERIDYNTQKQIRDEAAAGALRRHHRRLLRRRL GRRRRLRQRYMERKWTQLQNYRNPAGTNAQACLAVALKELNYPVIDITWKMQQFPD "	
complement(join(5163..5279,7637..7845,7935..8050,8149..8210))	
/gene="OSJNBA0051H17.2"	
complement(join(5163..5279,7637..7845,7935..8050,8149..8210))	
/gene="OSJNBA0051H17.2"	
/note="transposase-like protein"	
join(9375..9530,10947..11081)	
/gene="OSJNBA0051H17.3"	
join(9375..9530,10947..11081)	
/gene="OSJNBA0051H17.3"	
/note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard"	
11904..12794	
/gene="OSJNBA0051H17.4"	
11904..12794	
/gene="OSJNBA0051H17.4"	
/note="probably inactive due to 5' exon missing in CDS pseudogene, TNP2-like transposon protein"	
pseudo	
13002..15220	
/gene="OSJNBA0051H17.5"	
13002..15220	
/gene="OSJNBA0051H17.5"	
/note="probably inactive due to including stop codon(s) in CDS pseudogene, Rim2 protein"	
pseudo	
15984..16427	
/gene="OSJNBA0051H17.6"	
<15984..>16427	
/gene="OSJNBA0051H17.6"	
/note="start and end point are not identified"	
15984..16427	
/gene="OSJNBA0051H17.6"	
/note="predicted by GlimmerM etc."	
/codon_start=1	
/product="hypothetical protein"	
/protein_id="BAD88089.1"	
/db_xref="GI:57900494"	
translation="MRPRVLHQHLEVPFPTISASPPPLPRAFPSPSSSLAVVSRSLH PPSPPTPPHLMGERRRCLSLSLSLFLQHPRRRLVTDATDIDAADPDSASR RCRPLRRGRRRRRHPPHAAAPASSPPRRHATPLPLPPRRHCT"	
17301..20737	
/gene="OSJNBA0051H17.7"	
17301..20737	
/gene="OSJNBA0051H17.7"	
/note="probably inactive due to including stop codon(s) in CDS pseudogene, putative hydroxyproline-rich glycoprotein"	
pseudo	
complement(join(21536..21593,21725..21819))	
/gene="OSJNBA0051H17.8"	

misc_feature	complement(join(21536..21593,21725..21819)) /gene="OSJNBa0051H17.8" /note="hypothetical ORF predicted by RiceHMM this category is not included in IRGSP standard" 26939..29186 /gene="OSJNBa0051H17.9" 26939..29186 /gene="OSJNBa0051H17.9" /note="probably inactive due to including stop codon(s) in CDS pseudogene, P450" pseudo	Db	94767	GGTTCC-----GCCAGCCCTACCCGTGC---TCCCGC	94796
gene	complement(33543..33875) /gene="OSJNBa0051H17.10" complement(33543..33875) /gene="OSJNBa0051H17.10" /note="hypothetical ORF predicted by GlimmerM this category is not included in IRGSP standard" 35632..36027 /gene="OSJNBa0051H17.11" join(35632..35645,35814..36027) /gene="OSJNBa0051H17.11" /note="start and end point are not identified" join(35632..35645,35814..36027) /gene="OSJNBa0051H17.11" /note="predicted by GENSCAN etc." /codon_start=1 /product="hypothetical protein" /protein_id="BAB90700.1" /db_xref="GI:20161785" /translation="MGRRDAGGGRAGWPTACRRGGPGGLEDSQAARPPGRGGAVLGG GAAGVPGGLVAGFICDEFVIRICVFDLCVM" complement(41636..42103) /gene="OSJNBa0051H17.12" complement(41636..42103) /gene="OSJNBa0051H17.12" /note="retroelement-like protein" complement(42773..43616) /gene="OSJNBa0051H17.13" complement(join(42773..42832,43509..43616)) /note="start and end point are not identified" complement(join(42773..42832,43509..43616)) /gene="OSJNBa0051H17.13" /note="predicted by GeneMark.hmm etc." /codon_start=1	Qy	91	GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet	110
misc_feature		Db	94797	TCCCTCGCGCGG-----TTCGAGACGCCCGCGCACCCCGGCGCGC	94835
gene		Qy	111	GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySer	130
misc_feature		Db	94836	CGCAGCCCTCC-----GGCGCGCGCGCGCGCGAGCAGCGCA-GAACTGGACGGG	94888
gene		Qy	131	GlnGlyLysPro-----LeuHisArgArgArgSerGlnSer	143
gene		Db	94889	GGGCGGGTTCGCCGCCGCCGCCGCCATCCACCCGAGACGTCCTCCCTCTCT	94947
gene		Qy	144	ArgGluProGlyHisArgSerProSerArgGluProSerAla--AspGlyArgProSerG	163
misc_feature		Db	94948	CGGAGCGCGGAGCGAGGACACCCCGAGACCGCGCGCGCGCGCGCGCGCGCGGCG	95007
gene		Qy	163	luSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnV	183
gene		Db	95008	CCGACGCCACGCGG-----GTGTTCTGCTCGGCGCACCAACA	95043
gene		Qy	183	alaSerProAspValLeuArgAlaIleArgArgPheLeuHisSerTyrArgSerSerA	203
gene		Db	95044	TCAGCGTCGAGGACGTCAACGCCGCATCTCCGGTTCCTCCGCCACTTCCGC	95100
gene		Qy	203	laHisAspLeu-----AsnSerLysTyrIleGlnIleG	215
gene		Db	95101	CGCGGACGCCCGCGCGTCGACCGGTCATGAGGAGGCGGAGTACATCGCGCCATCC	95160
gene		Qy	215	luGluThrValGluArgGlu---GluAspThrLeuAsnIleAspMetSerAspIleTyrA	234
gene		Db	95161	ACCGCATCTCTGAGCTCGAGGCGCGGAGTCTCGACGTCAACGCCACCGAGCTGTTCG	95220
gene		Qy	234	sPheAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuL	254
gene		Db	95221	ACCAGACCCGACCTCTACGGCAAGATGGTCAGGTACCGCTCGAGGTGCTCGCCATCT	95280
gene		Qy	254	euAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleG	274
gene		Db	95281	TTGACATCGTGTCTATGGACCTCGTCGCGCGCATCGAGCCCTCTTCGAGAGCACATCC	95340
gene		Qy	274	luAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer----	292
gene		Db	95341	AGACCAGGATCTACAACCTCAAGTCTCGTTTGTCTTGAGGAATCTCAACCCCTTC-TGGT	95399
gene		Qy	292	-----	292
gene		Db	95400	GATTGCGCACTGCTCTGCTCTTTAGCTACTTTGTGTCGCGCAATCTTTTATANGATTTC	95459
gene		Qy	293	-----AspIleAspLysLeuValSerValLysGlyMetValIleA	306
gene		Db	95460	TGATTTTGGTTCGTTGTCAGATATCGAGAGATGTGTTCATCAGAGGTATGATTAATTC	95519
gene		Qy	306	rgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly-	325
gene		Db	95520	GGTGCACTCGGTGATTCAGAGCTCAAGAGGCTGTGTTCCGCTGCTGTTGTTGGGT	95579
gene		Qy	326	--HisSerProProLeuVal-----	331
gene		Db	95580	TCTACTCTGAGCCTGTAATGGTTGTCGAGGTATATATATATATATATATATATATATAT	95639
gene		Qy	331	-----	331
gene		Db	95640	CTTTTTCACACTGCGTATTCATATGCTCTGTTGTTTAAAGTGAACCTTCAGGTGATCAATTA	95699
gene		Qy	331	-----	331
gene		Db	95700	ATCAGATTGCGACCTTAATTTTGTCTCATAGAGATTACATATATGACAACTCATTTGTTAA	95759
gene		Qy	331	-----	331
gene		Db	95760	CTTCATTTGTACAAATCAITCTTTATCATGAAGATGATAAATTCAGCTTTCATTTTGACAT	95819

QY 331 ----- 331
Db 95820 GTTATATGTCACTAATTATCAATTATGTGCAATTTCTCTTTATTTGTTGTTGCAAAAACCTTT 95879
QY 331 ----- 331
Db 95880 TCATTTTGTGGATACACAACATTTAGGAATTTTATTATCATACCACATTTGCTGAGGA 95939
QY 331 ----- 331
Db 95940 ACTGATAAATAGTACTAATTAATGAGTAGCGCAAAATCAAAATCCATACCTACTGCTATTGTGCA 95999
QY 331 ----- 331
Db 96000 ATCCAAGACTCCGACTTACTGTCTGTCTGTATAAAGTAACACATTTTGCACCTGTGGTAGTAAT 96059
QY 331 ----- 331
Db 96060 TGTAAGTAAACATGCTTATGTTCTTAACTTACATTATATTCTGGTGGTTAATGTTGA 96119
QY 331 ----- 331
Db 96120 AGAGATTGTCATGCAGATCTTATGAAAAACATTAAGTGGCCTAGTGGAGTAATGACTA 96179
QY 332 ----- ThrVal - VallysGlyArg 337
Db 96180 AAATCCATCAGTTCTGTTGTTTATCATGTAACATAATTCACCTATTCTTTAGGGAGG 96239
QY 338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db 96240 GTTACTGCGCCGACATCTGTGAGAAAGAACAAATGTAAGCTACAACTCTATGACTCTT 96299
QY 358 IleHisAsnArgCysThr ----- 363
Db 96300 GTGCATAACCGATGCGAG - GTAGCTATATCCTACTTGTAAACCCCTTTTCTTTATAGGATT 96358
QY 364 ----- PheAla 365
Db 96359 TTTGACAGATATGTAGTCTGACCTTTGGATGGAAACATTTTATTTGTAACTAGGTTGCA 96418
QY 366 AsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGluGluThrPro 385
Db 96419 GATAGCAGATCATAAAGTTGAGGAAACACACAGATGAGATACCAAGAGTGGCACTCCA 96478
QY 386 HisThrValSerMetCysLeuThrAsnThrMetValAspAlaValLysProGlyAspArg 405
Db 96479 CATACTGTTAGTGTCTTGATGCACGATAGCTTTGTAGATGCTGGAAGCCCTGGAGACAGG 96538
QY 406 IleGluVal ----- 408
Db 96539 GTTGAGGTTAGCAGTGAATCTTTGTTGGTCAATTAACAAGAAATAATTTTGTGTTG 96598
QY 409 ----- ThrGlyVal 411
Db 96599 GTTCTTATTACATCGTAAAAATCTTGATTTTCTTTGATTATAGATACTGGATA 96658
QY 412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg ----- 427
Db 96659 TACAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGCAGTGAAGTCGATTTTCAAG 96718
QY 427 ----- 427
Db 96719 GTAGAGGATGTGAATGTTATCCTACTTGTCTATCTTAAATCTCTTCTGCACAGAAAA 96778
QY 428 ----- AlaLeuTyrlsThrTyrlle 434
Db 96779 AAGTTTAAATTTATCTACTCAAGCATCGTCAACAGTTTCTTTAAT - CAGACGTACATT 96837
QY 435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db 96838 GATTGCGCTTCACATAAAGAACAGACAGCAAGTCTAGACTTCTATGTTGAGGACTCCATGGAA 96897

QY 455 MetAspLysGluAsn ----- AspMetTyrlAla 463
Db 96898 ACTGATAACCCCAATGTCTAACAGACAACACTGAAAGATGATTTTCTCAGAGATAAGGTATCC 96957
QY 464 GlyTyrlHisGluSerAspThrSerGluAla ----- Ala 474
Db 96958 CTCAATCACCTTACTCTCTCTGGGATA - GCACCTAAGTGTAATGTGTTTAAACACACTGCTCCTT 97016
QY 475 AsnGluAlaLysIleGlnLysLeuLysLeuSerLysLeuProGlyLysLeuTyrlAspArg 494
Db 97017 TCTGTATTGCAAGTTGAGAAATTAAGAGATTGCAAAAGTTGCCAGATATATATATGACAGA 97076
QY 495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLeuLysGly ----- 512
Db 97077 TTAATAGTCAATTGGCTCCAAACATATGGAGCTGGACAGCGTTTAAAGGGGCGCTCCTT 97136
QY 512 ----- 512
Db 97137 TGCCAGGTAAGTATCATCTATCTGTTTGAATGTTTGTGCACCTAATAGCCTAGTATAACAAG 97196
QY 513 ----- LeuLeuCys - GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer 530
Db 97197 CTCTTACTGTTGTGTCAGCTTTTGGTGGAAATGCTTTTGAGGCTTCTCTGAGAGTAG 97256
QY 530 rPheArgGlyAspIleAsnValLeuValGlyAspProGlyThrSerLysSerGlnLe 550
Db 97257 TTTCCGAGCCACATCAATATTTGCTTGTGTTGATCTCTGGAACAGTAATATCCAGCT 97316
QY 550 uLeuGlnTyrlValHisLysIleAlaProArgGlyLysLeuThrSerGlyArgGlySer 570
Db 97317 TCTCAATACATGACAAACCTGCTCTCGTGGCACTTATACAAAGTGCAGAGGAAGTTC 97376
QY 570 rAlaValGlyLeuThrAlaTyrlValThrLysAspProGluThrArgGluThr ----- 587
Db 97377 AGCTGTTGGCCTTACTGTCATAGTTTACCAAGGATCCTGAAACTGCTGAAACTGTAAGTCT 97436
QY 587 ----- 587
Db 97437 ACATGAATCCTTTATTTTCGCAACTTAATTTAACTCGTTTAACTGATTAATCTGATTGATGA 97496
QY 588 ----- ValLeuGluSerGlyAlaLeuValLeuValLeuSerAspArg 599
Db 97497 CTTCAACCATGACATCTCATTTCCAGTTCTTGGAGATGAGCGCTTGTGTTGAGTGACAA 97556
QY 599 gGlyLysCysLysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHi 619
Db 97557 AGGTGTTGTTGTTATGATGAATTTGATAAGATGTTCTGATAATGCTCGAAGCATGTTACA 97616
QY 619 sGlu ----- 620
Db 97617 TGAGTCTCTTGTAAATGCTGCGCCCAAGATTTGTATCACAGTACTATTATTAGTCAATG 97676
QY 621 ----- ValMetGluGlnGlnThrValSerValAlaL 631
Db 97677 CTAAACTGTCCTTTTCAAAATATTGCAAGTGATGGAACACACAGACTGTCTCCATTGCCA 97736
QY 631 ysGlyGlyLysIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProS 651
Db 97737 AGCGTGGAAATAATTGTCATCTTTAAATGCAGAACATCAGTTCTAGCATGTGCAAAATCCTTA 97796
QY 651 erGlySerArgTyrlAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThrL 671
Db 97797 CTGAATCAGTTATTAATCCAAGGCTCTCTGTGATTGACAAATATCCATCTTCTCTCAACAC 97856
QY 671 euLeuSer ----- 673
Db 97857 TGTCTTCTAGGTGAGTGAATGCAATGGAATGAGCAATCTTCTTCCAGGACAAACAGATCA 97916
QY 674 ----- ArgPheAspLeu 677
Db 97917 ATAATAGAAGTGTTTTAACTCTCTGATTGTTTATTTTATTTTCAAAATTCAGGTTGACCTC 97976
QY 678 IleTyrlLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeu 697

DB:	2	Gaps:	14
US-10-768-511-8 (1-901) x AC116551 (1-153751)			
QY	179	TrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsn	198
DB	7589	TTGGGGTACACAGATTAAAGTTGAAGATGTTAAAGAAAGATTTTAGAAAATTTATATTAAT	7530
QY	199	Tyr-----ArgSerSerAlaHisAspLeuAsnSerLys-----	209
DB	7529	TTCCCAACCAAAAGAAAGAAAGATTCAATATATAATGACATATAATAATAATAAT	7470
QY	209	-----	209
DB	7469	AATAATAACAATAACAATAGTAGAAGAAATGAATGATGACGATGATTATTATTATGATGAA	7410
QY	210	-----TyrIleGlnIleIle	214
DB	7409	GAACAAGATAAAATTAATTAATAAAACCAGAAAAGTTCAATATTTATATAGAGATTATTATTA	7350
QY	215	GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp	234
DB	7349	AAAGTATTAAATGAACAACTAAATAATGTCAATTTAAATATAAATATGAATAATTTAGTACAA	7290
QY	235	HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu	254
DB	7289	TTTGATATGGAATTTATTTACAAATGGGTATCATTTCCAAATGAATGATACCAATTTATTA	7230
QY	255	AspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHis-----	272
DB	7229	GATGAGGAATTAATTAATTTATAGAGAATTAATTTCCAAATACACCAAGAAATGGTGAT	7170
QY	273	-----IleGluAlaArg	276
DB	7169	GATGAAGATGATGACGATGATGATGAAGATGAAGATGATTAAGAATTGAATACAT	7110
QY	277	ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLys	296
DB	7109	CCATTCATTTTAAAGAAAACACCAATGAGAGATTTAAATCCATCAGATATGATAAG	7050
QY	297	LeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGly	316
DB	7049	ATTATATCAATTAGTGGTTAATTTATTAAGTAGTTCAATTTATATACAGAGATTAAACAA	6990
QY	317	AlaPhePheLysCysLeuValCysGly-----HisSerProProLeuValThrVal	333
DB	6989	GGGTTTTTTATGTGTGTCAGTTTGTGAAGCAACTTTTTCATGCA-----AATGTT	6942
QY	334	ValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsn	353
DB	6941	GAGAGGGTAAATTCAGAGCCATCAGATGTTCA-----AATGTAAATCAAGCAA	6888
QY	354	AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln	373
DB	6887	TCATTATCAATCATTCATTAATCGTTGTTGTTGGTGATAAACAGTATATAAACTTCAA	6828
QY	374	GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr	393
DB	6827	GAACAACCAAGATCAATTCAGAGGGTGAACACCAATACCGTCGCATTTATTGCTAT	6768
QY	394	AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys	413
DB	6767	GGTGATTAAATGATATTGCAAAACCAAGGTGATCGTGTGAGTAACTGGTGATTATTA	6708
QY	414	AlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyr	433
DB	6707	GCAAGTCCAATGAGAGCTGGTTCAAT---AGATCACTTAGAAGTATTTATATAAACTTAT	6651
QY	434	IleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMet	453
DB	6650	ATTGATATTTTACATATAAAAGACAGATAAAGGTAATATATGATGATGATGATGAT	6591
QY	454	GluMetAsp-----LysGluAsnAspMetTyrAla	463
DB	6590	CATGATGATAATACTGCTGGTGGCACAGGTACAGGTAAAGAAACAATAATTTA	6531
QY	464	GlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLys	483
DB	6530	GATTTTGAAGATTTAGATGAATTTGATTTATCTGAAGAA---AAAGAAAGAGAGATTATA	6474
QY	484	GluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIle	503
DB	6473	GAAATTATCAAAAAAACAGACATTTTATGATATTGTTAACTAAATCAATAGCACCAATAAT	6414
QY	504	TrpGluLeuGluAspIleLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLys	523
DB	6413	TGGGAATTAGAGATATTAAAAAAGGTATCTTTGTCATTTATTTGGTGTTCAAAAAG	6354
QY	524	LysIleProSer---GlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAsp	542
DB	6353	AGTTATCAAGATTATGGTGGAAAGTTTAGAGGTGATATAAACAATTTATTGTGTGTAT	6294
QY	543	ProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIle	562
DB	6293	CCTGTACTTCGAAATCTCAATTAATCTCGTATGTTTCAATAAATTCACCCCGTGTATC	6234
QY	563	TyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspPro	582
DB	6233	TATACAAGTGTGAAGGTAGTAGTCAGTTGTTTGCAGCGCATATATCAACCAAGATCCC	6174
QY	583	GluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCys	602
DB	6173	GATACTAAGAGACCGTTTATAGAGTGGTGTGATGTTGTTCTCTGATGAAGGTGTTGT	6114
QY	603	CysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMet	622
DB	6113	TGTTATGATGATTTGATAAATGAATGATCAAACTCGTTCAATCTTCATGAAGTTATG	6054
QY	623	GluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThr	642
DB	6053	GAACAGCAACACAGTATCAATTCGAAAGCGCGCATTTATCTGCACGTTGTAATCGCAACC	5994
QY	643	SerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIle	662
DB	5993	TCGATCTTGCACTGCAAAATCCAGTGGTTCCAGTTATATGCCCAAAATATCAGTGGTG	5934
QY	663	AspAsnIleGlnLeuProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeu	682
DB	5933	GAGATATCCAAATGCTCCCAACCTTATATCGAGATTCGATTTAATTTATTAGTATTG	5874
QY	683	AspLysProAspGluAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyr	702
DB	5873	GATAAAGCCAAATGAAAGAGATGATCGTCAATTTGGCTCGTCAITTAGTATCGATTTGG	5814
QY	703	GluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIle	722
DB	5813	GATGAACACCGGTTTCT---CATTTCAACATCCCCCAAGNAACCTTCACAACTATAT	5757
QY	723	ThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIle	742
DB	5756	CAATACGCTAGAAAACATATTAATCCAAAGTTAACTGATGATTCGCAAAATGTTAGTT	5697
QY	743	AsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIle	762
DB	5696	CAAGGTTATTTAGAGATGAGAAGTATG-----GGTAGTAGTAAAAAACAATC	5649
QY	763	ThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMet	782
DB	5648	CTGCGCACCAAGACAATTTGGAATCACTTATTCGTATCGTGAAGCTCATGCTGAAT	5589
QY	783	ArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspVal	802
DB	5588	CGTTTCTCTGAATTTGTTGACCAATTAATGATGTTGAAGAGCAATTCGTTTAAATAGTC	5529
QY	803	AlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThr	822

[illegible]


```
Qy 667 uProProThrLeuLeuSerArgPheAspLeuLeuTyrLeuMetLeuAspLysProAspG1 687
Db 2653 TCCGCTACTCTCTATCTAGATTGATTTGGTGATTATTAATATAGACCGCGTGGATGA 2712
Qy 687 uClnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn-----Ty 705
Db 2713 AACGCTGGATAGGAAGCTCGCAAAATCATATTGTTCTATGTACATGGAAGATACCTCTGA 2772
Qy 705 rGluValSerLysGlnAspAlaLeuAspLeuThrLeuThrAlaTyrIleThrTyrAl 725
Db 2773 ACATGCTACAGATATGGAGCTTTAGTGTGGAATTTTGACAGCTATATCACAATATGC 2832
Qy 725 aArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTy 745
Db 2833 TAGAATAATAATTAATCTCTGTCATTTAGTGAAGGCGAGCAAGGAACCTAGTGAATGCTTA 2892
Qy 745 rValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaTh 765
Db 2893 TGTGCTATGCGAAAGCTTGGGGAGATGTGCGAGCATCTGAGAAAGAAATTTACTGCTAC 2952
Qy 765 rProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSe 785
Db 2953 AACAGACACACTGGAAATCATGATTCGTCTATCTGAGGCCCATGCGAAATGCATCTTCG 3012
Qy 785 rGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuG1 805
Db 3013 CAACGTTGTTGAGTAGGAGATGTTCTTGAAGCTGCTGCTTGATTAAGACCGCTATTAA 3072
Qy 805 nClnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrGlyVa 825
Db 3073 GGATTTATGCACTGATCTCTGCAACAGGAAGAAATTTCTTTGACCTAAATTTAT-----GT 3126
Qy 825 IserAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaLeuArgGluLeuIleAl 845
Db 3127 TAATGAAGAGACACTTGTGTACCTGACGATATGTGCAAGAACTGGCGCAATTAATTC 3186
Qy 845 aAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAsp1 865
Db 3187 AAATTTAACTGTGGTGGGAAAACT-----ATGTTGGTCAGTCAATTTGTAACCGGTT 3240
Qy 865 eArgSerGlnSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLe 885
Db 3241 TCGTGACGACAGTTCTACTCGCTTAGATCTTCTGATTTGGAAGCTTGTCTGGGTGCTTT 3300
Qy 885 uGln 886
Db 3301 TCGT 3304

RESULT 9
AC149637
LOCUS AC149637 116616 bp DNA linear PLN 20-MAY-2005
DEFINITION Medicago truncatula clone mth2-181h2, complete sequence.
ACCESSION AC149637
VERSION AC149637.11 GI:66351814
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 116616)
AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.
and Roe B.A.
TITLE Medicago truncatula BAC Clone mth2-181h2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 116616)
AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.
and Roe B.A.
TITLE Direct Submission
Submitted (17-JUN-2004) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
```

```
REFERENCE 3 (bases 1 to 116616)
AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.
and Roe B.A.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2005) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 116616)
AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.
and Roe B.A.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-2005) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On May 20, 2005 this sequence version replaced gi:62988494.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR
-----
FEATURES
source
1..116616
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-181h2"
/clone_lib="Medicago truncatula BAC library H2"
ORIGIN
Alignment Scores:
Pred. No.: 4.8e-59 Length: 116616
Score: 1858.50 Matches: 505
Percent Similarity: 37.18% Conservative: 140
Best Local Similarity: 29.11% Mismatches: 182
Query Match: 40.32% Indels: 912
DB: 15 Gaps: 26
US-10-768-511-8 (1-901) x AC149637 (1-116616)
Qy 12 ValSerSerProTyrProSerGlnSer----- 20
Db 42167 GTTTCATCTCCATCTTATCTAATTCATTTCTAATTTCTGAAACCCCT 42226
Qy 21 -----GluGlyValSerThrProLeuProGlnValThrSerProSerPheAsp 36
Db 42227 ATTTTTCACCTTTTCAGGTCCTTCTTCCACCGACAACTCCAATTCAGATCGCAAC 42286
Qy 37 AsnAlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaVal 56
Db 42287 ACCTACTCTCTCCCTC-----GGCCGCCGT----- 42310
Qy 57 ArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSerLeu 76
Db 42311 CGCCCGCGTCAATCATCTCACTCCATCTGGAACCTCTCAACCGACAGATCTCGC----- 42364
Qy 77 GlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrPro 96
Db 42364 ----- 42364
Qy 97 ValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHis 116
Db 42365 -----TTCGCTTCTCTCCGATGCAACTCCA-----ACTCCCTCCCTCC 42403
Qy 117 ArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHis 136
Db 42404 CGT-----CGAGGTGGCGGAATCCCTCCCGC 42430
Qy 137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
Db 42431 CGCCGTGGACCGCCACACCAACTTCCACTCCGAGACTTTCCTATGCTTCTGGAAGGC 42490
Qy 157 AlaAspGlyArgProSerGluSerAlaGluProAspThrLeuGlyGlyGlyTyrAla 176
```

```
Db 42491 GGTGACGGTTATGACATGGATGATGTCGTCTCT-----ACT 42526
Qy 177 TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu 196
Db 42527 TACGTTTGGGGTACTAAATGACGTCGAGGATGTTAAACGATGCGGATTCAGAGGTTTTTG 42586
Qy 197 HisAsnTyrArgSerSerAlaHis----- 204
Db 42587 AAGCAATTTAGAGAAAATCTGCTCTCAAGCGGTGATGATGATTTGGATATGGAATTTG 42646
Qy 205 AspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThr 224
Db 42647 GATATTGAAGGAAGTATGAGAAATGATCAACAGAGTTATTGAATTCGAAGGTGAGTCT 42706
Qy 225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244
Db 42707 ATTGATGTTGATGACAGTCATGTTGATCAGCATCAGCAATTGTATAGTAAGATGGTG 42766
Qy 245 ArgTyrProLeuAspIleProLeuLeuAspThrGluCysGlnGluValAlaLysThrSer 264
Db 42767 AGGTATCTCTCGAGGTGCTGCCAATTTTGATATGGTTTGGATGAACATGGTCACTAGG 42826
Qy 265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284
Db 42827 ATGAATCTATGTTTGAGAGCATGTTCAAACTAGAGATTTTAACTTTAAGAGCTCAACT 42886
Qy 285 HisMetArgGluLeuAsnProSer----- 292
Db 42887 TCAATGAGGAATTTGAATCTCTCGTGAGTTTTTTGTTTTTTGAAATATTATTGTTATGA 42946
Qy 292 ----- 292
Db 42947 ATTCAATTTGTGCTTTGATTTTAGCAATTAGATATGATATGTAATGATGTTCTTGAT 43006
Qy 293 -----AspIleAspLysLeuValSerValLysGlyMetValIleArgCysSer 308
Db 43007 GTTTGTTTGAAGATATTGAGAGGATGTTGTCGATGAAGGNAATGATTATTGCGTGCAGC 43066
Qy 309 SerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerPro 328
Db 43067 TCGATTATACCAAAATCAGGAAGCTATATTTCAGGTGCTAGTATGTTGTTCTGCTCT 43126
Qy 329 ProLeuValThrValLys----- 335
Db 43127 GACCCGGTTCGGTGGAGAGGTAATTACAGAGTCATTTCTGGTTATTTATGTAGTTAA 43186
Qy 336 -----GlyArgVal 338
Db 43187 CTGTTGTGTTATATCTGAATGTAATATTTTGTCTGGCTATTTTGATTTTTAGGACGAATC 43246
Qy 339 GluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIle 358
Db 43247 GCCGAACCTACTGTTTGTCTGAGGGAAGAGTGTCAAGTCAGGAATTCGATGACACTGTT 43306
Qy 359 HisAsnArgCysThr----- 363
Db 43307 CACAACCCGATGCAA -GTAAGTCTGCTTATTGTTGTTAAAGAAGTGATTTTTTTTAGTTGA 43365
Qy 363 ----- 363
Db 43366 TTATTTGGNAITGGAGTACTATATTTTGTGCTGTTGGATAGTGTGTAACATGTTATGCAATC 43425
Qy 363 ----- 363
Db 43426 ATATTTGATATTGAAGGGCTATATTCTGAAGCAAGCTCTTTTAAACCTTTTTCCTTTAA 43485
Qy 363 ----- 363
Db 43486 CTCACATGCTATGTATATAGATAGTAGTCAGTTCTGTTCTTCTGATTTTTTGGTTGTG 43545
Qy 363 ----- 363
Db 43546 AGCAATCAGTGGCAGCAGTGAATTTTCTGTCATAGTATAAAGATAGGCAAAATAGCTGGATA 43605
```

```
Qy 363 ----- 363
Db 43606 TGSAAAGTTAACTTTCTGAACCTGGTCTACTACTAAATATTATTTCCATGCAACACGACTAAT 43665
Qy 363 ----- 363
Db 43666 AACCTGCTGGTGGTTTTAGTGTAAAAATTGACCAATGCTCTGGCTAAAAATTGGGTTAATTTTT 43725
Qy 363 ----- 363
Db 43726 TACAAAAATCAATCTGATCTAGAAAGTGAGGAACCTGTGCTTTATTATTGTCAGAGGTTTAC 43785
Qy 363 ----- 363
Db 43786 GATAAAAAACAGTTTGAATAAATGATTATTTTGAACCTGTACTGGGTATATTTTAGTGATATGT 43845
Qy 363 ----- 363
Db 43846 TGGATGCTTGGCATGGTAAATTATCCCTGTTTTTAAACAATGTGTTTGGTCATGTGACTA 43905
Qy 363 ----- 363
Db 43906 TTTGCGGTAAAGTTTTTTCACCTTGGATTTTCAGTTCATAAACCTTGAAACTGCAATTAATTAAT 43965
Qy 364 -----PheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db 43966 ATCTTCTTAATTTTCAGGTTTTTCTGATAAACAAATTTGTAGGCTTCAGGAGACACCTGAT 44025
Qy 378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db 44026 GAGATACCTGAAGGAGGAACACCTCACACTGTTAGCTTGTCTATTGTCATGACAACTGGTG 44085
Qy 398 AspAlaValLysProGlyAspArgIleGlu----- 407
Db 44086 GATACTGAAAGCCAGGTCAGAGATTGAGGTGAGACTATACCATGAGATACTCTGTATAA 44145
Qy 408 -----ValThr 409
Db 44146 ATAGATATTCATTATTTTGTTCCTCAATCGTCCCATCAGTGTGTTTGTGTATCATCAGTGTACT 44205
Qy 410 GlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeu 429
Db 44206 GGTATATACAGGCTATGAGTGTCAAGGTGGACCACTCAGAGAACCTGTTAAGTCATGT 44255
Qy 430 TyrLys----- 431
Db 44256 TTCAAAGTATGGGTTACCTTCTCAAACCTCAATTTGCAATTTAGAGCTTCTCTTTTCATTTT 44325
Qy 432 -----ThrTyrIleAspCys 436
Db 44326 TCATCTTTGCTCTAAATTGACATTTGTTTCTTTTGTTCATGTTGTAGACGTATATTGATTTGT 44385
Qy 437 ValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAsp 456
Db 44386 CTTTCATATAAAAGACATCAAAAGTCCCGATGCTAGTTAGAGTGTATGAGAGGTTGAC 44445
Qy 457 LysGlu----- 458
Db 44446 AATGCCCAGGCGAGAAATCCAGAAAGTTTTTATTGATGAAGAAAAGGTACATTTTGTGA 44505
Qy 459 -----AsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAla 475
Db 44506 CATCTCTTCAATGAATATCTGCTGTTTATTATCCCGCTAAACTGCA---TGCTTTGTCT 44562
Qy 476 GluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeu 495
Db 44563 GAGTGTCAAGTGGCTAAACTGAAGAGCTCTCAAAAACCAACGATATCTATGAAGAGGCTG 44622
Qy 496 SerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCys 515
Db 44623 ACAAAAGTCACTGGCGCCAAACATCTGGGAGCTAGATGATGTAAGAAAGGCTCTTCTTTGTC 44682
```

```
QY 515 ----- 515
Db 44683 CAGGTAAGACATCTTTCTAGCTAGTCTTTTCATTTGTTGTTTACAAATATCTGAGGATG 44742
QY 516 -----GlnLeuPheGlyGlyLysAlaLysValLeuProSerGlyAla 529
Db 44743 CTGGTTTTATTTCTTCAGCTTTTGGGGCAATGCTTTGAAAGTTGGCATCTGGTCT 44802
QY 530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db 44803 AGCTTTCTGGTGATATAAATATCTCTTGTGTGTATCTCTGATCAAGCAAGTCGCAG 44862
QY 550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db 44863 CTACTTCAGTACATACAAAATCTCTCCACGTCGCAITTTACACAGTGAAGGGGAAGC 44922
QY 570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr 587
Db 44923 TCAGCAGTTGGATTGACTGCTTACGTGGCCAAAGATCTGAAACAGCGGAACACTGTATAG 44982
QY 587 ----- 587
Db 44983 ACCAAGTCATGATCTTTTAGTGTCTATAGTACAGAAATGCAATGGAGTCTAAATTTTT 45042
QY 588 -----ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCys 602
Db 45043 CTGCTAAATCTTCAGGTTCTTCAGAGTGTGTGCCCTAGTTTTCAGCGCAGCGAGGTATCTGC 45102
QY 603 CysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db 45103 TGCATTGTGAATTTGACAAAATGCTGCAATGCAAGGAGCATGTTACACGAGTGTGT 45162
QY 620 ----- 620
Db 45163 TTCATTTTGTGACGATGATTTCTGCAGTCTCTATCAGCTGTTTTTTTCTACAGAACTT 45222
QY 620 ----- 620
Db 45223 TGCATGGATTGAATAATGGTAAATGTTGGCACTAGTTAAGGAGACTAAATTTAGAAATTTT 45282
QY 620 ----- 620
Db 45283 GCATTTGMAATTTGTTAAAGAGTGATGTTCAACTTTTTTAAAGTCAAAAGTTGTGTGT 45342
QY 620 ----- 620
Db 45343 TTCAAGATAAAATTTCTATTTTCTTTTACTTTCTTAACTAGTGCCCCGGGGGCACCTAGTTAAC 45402
QY 620 ----- 620
Db 45403 ATGACCCCTTGAAAATATATACACTATGATTTTCGTCTTTATCATGTTAGCTTTGTACTT 45462
QY 621 -----ValMetGluGlnThrValSer 628
Db 45463 TTGAGTAACATAACACTTGTCTCTGTTTATTCAGGTCGATGGAACAACAACCTGTTTCA 45522
QY 629 ValAlaLysGlyGlyIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAla 648
Db 45523 ATAGCAAGGCGAGCATTTATGCTTCCCTCAATGCAAGGACCTCGGTGTGGCTTGCACA 45582
QY 649 AsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuPro 668
Db 45583 AACCTTAGTGGTFCAGATATATATCCCGCTGTCTGTCTATTTGACAAATACACCTTCTCT 45642
QY 669 ProThrLeuLeuSer 673
Db 45643 CCCACTCTTTGTGCCAGGTTAGTTGTTGGACATAAGCCGTGATACTAGGTTATGATCT 45702
QY 673 ----- 673
Db 45703 GGCTACGTATTTATGCAATTTCTCTCATGATGTCGATGTCGACGAATCAATTTTATACCTTGGGCCT 45762
QY 673 ----- 673
Db 45763 TACCTCTATTTTTTTTGGGTGTTTATTTGAACATAACACAAAATTTGAATCTGTGTGTGATT 45822
QY 673 ----- 673
Db 45823 CTCCTTATCAAACTAGTGTGGTAGGGTTGGCATAAGCAAAAATTTCTCTTTAAAAA 45882
QY 673 ----- 673
Db 45883 TAATGGAAGCTACTAGACAGATTTTTCGAAGTGTGTCTATTGATTTTAAGTTGATCTTTTTC 45942
QY 674 ---ArgPheAspIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgAr 692
Db 45943 TGGCAGATTTGATTTAATTTACTTAAATGCTTGAATAGGCTGATGAACAGACCCGACGAGC 46002
QY 692 gLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSer 708
Db 46003 ACTTGCCAAACATATTGTTTCCCTGCATATAAGGATTTATGAGTCTG-TTTTATAATCTT 46061
QY 708 ----- 708
Db 46062 GTAATCATCTTATTCAAGATTAAAGATCCAAACTTAAGTGTTTTTCATATCTCATCTATTAGAA 46121
QY 709 -----LysGlnAspAlaLeuAspLeuGlnThrLeuTh 719
Db 46122 TGATATGGCCATTAATTCGAGAACATTTGACAGCATGTTTGGACATTTCTTACACTAAC 46181
QY 719 rAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGl 739
Db 46182 TGATATGTTAGTATGCCCCGAAAGCACAATACACCCACAATCTATCAGATGAAGCTGCTGA 46241
QY 739 uAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerly 759
Db 46242 TGAGTTGATTACAGGCTATGTAAAAATAAGGGGTAGGGGAAAATTTCTCCGGAAGCAGCAA 46301
QY 759 sLysVal 761
Db 46302 AAAGT-GATACAAAGTGAATTAATTGATCTTGAATTAATGCTTATTTTTCATTTTTTTC 46360
QY 762 -----IleTh 763
Db 46361 ATGTTTATCTGTTGTGATGTTAATCTTGCCTATCTATCAATTAATTAACAGGTAATAAC 46420
QY 763 rAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetAr 783
Db 46421 GGCAACTCCAGGCGAGATTGAGAGTCTATTACGCTTAGTAGAAGCATTTGGCTCGGATTCG 46480
QY 783 gPheSerGluVal 787
Db 46481 CTTCCTCAGAAATC-GGTTAGTTGAGCTTGTGCTCTCTCTCTGATTTGTAGATAAAAAAT 46539
QY 788 -----ValGluLysValAspAlaAlaGluAl 796
Db 46540 AGCATCCAGAAATTAATTAAGTAACCTTGTTCACAGGTCGAAAACATGACGTGGTGAGGC 46599
QY 796 aValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db 46600 ATTTCCGCTCTCTGGAAGTTGCAATGCAATCTGCAATGGATACCAGAACTGGTAGTTA 46659
QY 811 -----HisAla 812
Db 46660 TCCAAACAGCCCTATCTCTATTGTTATTCACCAACAAAATAATATATATATATATATAT 46719
QY 813 -----ThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAla 828
Db 46720 TCCTTTTATCTTCACAGGAACATTTGACATGATCTTATACTACTGTTTTCGGCAAG 46779
QY 828 rGluArgIleArgArgAlaAsnLeuAlaAlaLeuArgGluLeuIleAlaAspLysI 848
Db 46780 CCAAGGATGAGCGGAGAAGTTTGAATCAAGACACCCCGCAACATGATCATGGAGAGAT 46839
QY 848 eSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGl 868
```

Dh 46840 CCAATTGGGGACGTTCC---ATCGGATTATTAGAGGTATTACCATCTTCTTCTAAACC 46896

Qy 868 nSerSerValAspVal----- 873

Dh 46897 ATTCAATAGATTATGTTTCAGTTTGAGACAATGATCTAATATAATTAAACAGATCAG 46956

Qy 873 ----- 873

Dh 46957 CAGATATTGGAGAGTTGAAGAAATCAGAGCCCTGCGACGAAATTCATCTTTAATGATGA 47016

Qy 873 ----- 873

Dh 47017 AGCCCTTTCCTTGAATATTTTACCAGTGTATATGCTTTTGCTCCATCCCGAGGATA 47076

Qy 874 -----SerLeuGlnAspIleLysAenAlaLeuGlySerLeuG1 886

Dh 47077 AGTGAACAAATGAGCTATTTCATGTTTAATGCGAGCTAAGGAGTCAGTTGCTACTCTTGC 47136

Qy 886 nGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900

Dh 47137 CAGTGAAGGATTGTTTCATGGCTGGTGCACAGTGTAAAAAGA 47179

RESULT 10

SCU14731
LOCUS SCU14731 5392 bp DNA linear PLN 31-JAN-1996
DEFINITION Saccharomyces cerevisiae Cdc54p (CDC54) gene, complete cds.
ACCESSION U14731
VERSION U14731.1 GI:608170
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 5392)
Whitebread, L.A. and Dalton, S.
AUTHORS Cdc54 belongs to the Cdc46/Mcm3 family of proteins which are
TITLE essential for initiation of eukaryotic DNA replication
JOURNAL Gene 155 (1), 113-117 (1995)
PUBMED 7698653
REFERENCE 2 (bases 1 to 5392)
Dalton, S.
AUTHORS Direct Submission
TITLE Submitted (15-SEP-1994) Stephen Dalton, Department of Gene
JOURNAL Regulation, Roche Institute of Molecular Biology, 340 Kingsland
Street, Nutley, NJ 07110-1199, USA
FEATURES
Location/Qualifiers
1..5392
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
1921..4722
/gene="CDC54"
1921..4722
/gene="CDC54"
/codon_start=1
/product="Cdc54p"
/protein_id="AAA86310.1"
/db_xref="GI:608171"
/translation="MSQQSSPTKEDNNSSPPVNPDSVPPQLSSPALFYSSSSSQG
DIYGRNNQNSQCGNTRAAIGSSPLNFPSSQRQNSDFQSQQRQIRSSASAG
RSYHSDRSDEALPTSSSLGRNQNRVHMRNDIHTSDLSSPRIVDFDTRSGVNT
LDTSSSAPPSSALSPFLRIWGTNVSIQECTNFRNLFMSFKYFKRLKIDREEFINN
TDCDELYIKQNLNRELGTSNLDARNLAKYQTEDLYHQLNYPQEVISIMDQTI
KDCVSLYIDLDLDEIETKFKYKVPYNNVSKGMRNLNPDIDKLNKGLVLR
STPVTIDMKVAFPKNCVDHTMAVEIDRGVQEPACRERIDCNEPNSLSLTHNRCSFA
DKVTLKQETPVPDQTPHSISLCVVDDELVDSCAGRIEVTGTFERSIPTRANSPO
RVLSKLYTYVDVHVKKVSKRLVDSTSTIEQELMNQNDHNEVEVRQITDQDLAK
IREVAAREDLVSLARSAPSIYELEDVKKGILLQLFQGTNKTFTKGRYRGDINILL
CGDPSTSQILQYVHKITPRGVTSKSGSAGLVTAITRDVDTKQVLVSAGLVS
DGVCCIDIEFQKSDSTRSLVHEVMEQQTISIAGAKIITLNARSLILASANPISGRY
NNPLVTENIDLPPLLSRFDLVLYLVDKDEKNRELAKHLTNLYLKDKEPHISODD
VLVPEFLTWLYSYAKEHITPIITEAAKTELVRAYVGMKMGDDSPSDEKRIATATROL
ESMIRLAENAHAKMKLNKVVLELVDQVEAVRLRSALKDYATDPKTKIDMNLVQTGKSV

source

gene

CDS

IOKRLQEDLSREIMNVLKQASDSMSFNELIKQINHSQDRVESDIDQALSRLQOED
KVIVLGGVRRSRVRLNRRV"

ORIGIN

Alignment Scores:
Pred. No.: 2,03e-59 Length: 5392
Score: 1833.00 Matches: 393
Percent Similarity: 59.36% Conservatives: 178
Best Local Similarity: 40.85% Mismatches: 281
Query Match: 39.77% Indels: 110
DB: 15 Gaps: 14

US-10-768-511-8 (1-901) x SCU14731 (1-5392)

Qy 2 GluAenAenAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
Db 1951 GAGGATAATAACTCC-----AGCTCCCCCGTTGTGCCTAAATCCTGAT 1992

Qy 22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAspSerPro 41
Db 1993 TCTGTTCCTCA-----CCACAGCTTCTTCCCCCAGCTCTATTATTATAGCTCCTCTTCA 2043

Qy 42 ValAlaGly----- 44

Dh 2044 TCACAGGTGATATCTATGTCGCAACATAGCCAGAACATTAAGTCAGGGAGGGAAC 2103

Qy 45 ---ArgArgAlaValArgGlnThrPro-----ThrSerAlaValArgArgGly 60
Db 2104 ATCAGAGCTGTATAGTGTCTCTCCACTAAATTTTCCATCTCTCTCCAAAGACAAAAT 2163

Qy 61 ArgGluThrAspSerAlaArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db 2164 TCCGATGTTTCCAAATCTCAAGGCAGACAGGCAGAAATTCGTTCTCTCGCAGCGCTTCT 2223

Qy 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db 2224 GGAAGTCTAGATATCATCTCTGATCTGAGAGTGTAGACGACCTGCTACTCTCTTCC--- 2280

Qy 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db 2280 ----- 2280

Qy 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
Db 2281 -----TCTTCTTTAGCGCGAATGCTCAAAACCGGTGTACACATGCGAAGA 2325

Qy 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db 2326 AATGATATTCATACATCTGATTATCTCTCTCTAGAGAATTTGGATTTTGATACTAGA 2385

Qy 161 -----ProSerGluSerAlaGlu 166
Db 2386 TCCGCGGTGAATACTTTTGGATACTTCTTCTTCTCTCTCTCCATCGAAGCCAGTGAA 2445

Qy 167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGlyThrAsnValAsnIlePro 186
Db 2446 CCC-----TTGAGAATAATTTGGGGTACCAACGTCATGATATCCAG 2484

Qy 187 AspValLeuArgAlaIleArgArgPheLeuHisAenTyrArg----- 200
Db 2485 GAATGTACAACTAATTTTCGTAATTTTGTGATGCTTTCAAGTATAAATTCGTAATAATA 2544

Qy 201 -----SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
Db 2545 TTGGATGAAGGGAGGAATTCATTAAACAATACTACCGACGAGAACTATATCTATATCAAG 2604

Qy 213 IleIleGluGluThrValGluArgGluAspThrLeuAsnIleAspMetSerAspIle 232
Db 2605 CAGCTTAAATGAATCAGAGAACTTGGTACCTCTAATTTAAACTTGGATGCTAGAAACTTA 2664

Qy 233 -----TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
Db 2665 CTTGCTTACAGCAACACAGAGACTTATACCATCAACTATTAAATATTATCTCTCAAGAGTG 2724

QY	251	IleProLeuLeuAspThrGluCysGlnGluValAlaThr-SerLeu-----	265
DB	2725	ATTTCTATCATGATCAAACTATCAAGGACTGTATGTTTCTTAAATAGTAGACAACAAT	2784
QY	266	-----LeuProThrPheGlu-----LysHisIleGluAlaArgProPheAenLeu	280
DB	2785	TTGGATTACGATTTAGATGAATAGAGACCAAGCTTTTATAAGTAGAGGCTTACAATGTA	2844
QY	281	LysAlaSerValHisMetArgGluLeuAenProSerAspIleAspLysLeuValSerVal	300
DB	2845	GGTCTCTGTTAAAGGTATCGCGAATTGATCCAAATGATATTGATTAATGTATAAACTT	2904
QY	301	LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys	320
DB	2905	AAGGGTCTTGCTCTCGATCAACTCCGGTATTCCCTGATATGAAGGTAGCGTGTTC	2964
QY	321	CysLeuValCysGlyHisSerProLeuValThrValValLysGlyArgValGluGlu	340
DB	2965	TGCAACGCTCGCATCATACA-----ATGGCAGTGGAAATTCATGATGAGGATTAT	3021
QY	341	ProThrArgCysGluLysProGluCysAlaAlaArgAenAlaMetSerLeuIleHisAen	360
DB	3022	CCCGCTAGGTGTGAACGATTGATTAATGAACCCAAATTCATGTCTCACTGATTCACA	3081
QY	361	ArgCysThrPheAlaAenLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro	380
DB	3082	AGGTGTTTCATTCCGCAGATAACAAGTCATTAAAGTTACAGGAAATCCAGACTTTG	3141
QY	381	GluGlyGluThrProHisThrValSerMetCysLeuTyrrAenThrMetValAspAlaVal	400
DB	3142	GATGGACAAACGCTCTACTCTATCTCATTAATGTGTTTACGATGAATAGTGGATCTT	3201
QY	401	LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly	420
DB	3202	AGGCGCGCGCATCGTATTGAGGTGACTGGCACGTTTCAAGTCCATCCCATTAGAGTA	3261
QY	421	ProAenGlnArgThrLeuArgAlaLeuTyrrLysThrTyrrIleAspCysValHisValLys	440
DB	3262	TCCAGGCCAACCGCTACTTAAAGTCGTTGTTATAAACATACGTCGATGGTGGTCCACG	3321
QY	441	LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp	460
DB	3322	AAAGTTTCAGATAGACAGTTTAGCGTCGATCTCTACTATTGNAACAAGAAATTAATG	3381
QY	461	MetTyrrAlaGlyTyrrHisGluSerAspThrSerGluAlaAlaAenGluAlaLysIleGln	480
DB	3382	AACAAGGTAGATCATAACGAGGTTCGAAGAAGTAAGACAAATTAATGATCAGGATTT	3441
QY	481	LysLeuLysGluLeuSerLysLeuProGlyIleTyrrAspArgLeuSerArgSerLeuAla	500
DB	3442	AAAAATTCGCGAGGTTGCGCGCAAGAGAAGATTTATACAGATTATTAGCCCCGCTT	3501
QY	501	ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuCysGlnLeuPheGlyGly	520
DB	3502	CCAAGTATTACGAGCTAGAGATGTCNAAGNAGGTATATTACTTCAGCTATTATTGG	3561
QY	521	LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAenValLeuLeuVal	540
DB	3562	ACGAATAAAACCTTTACCAAGGTTGGCGGTTTATAGAGTCACATAAATATTTTACT	3621
QY	541	GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrrValHisLysIleAlaProArg	560
DB	3622	GGGGAATCCTTCTACTTCCAAATCGCAAAATTTTGCAATACGTTCCAAANATTCCT	3681
QY	561	GlyIleTyrrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrrValLys	580
DB	3682	GGTGTGTATATCTCGGTAAAGTTTCATCTCGCGTTGGTTAACTGCTTATATACAAG	3741
QY	581	AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly	600
DB	3742	GATGTGCACACAAACACTGTTTGTGAAAGTGGTGCAATAGTATTGTCTGTATGGAGGT	3801

SOURCE	Saccharomyces cerevisiae (baker's yeast)	gene	
ORGANISM	Saccharomyces cerevisiae	CDS	
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
JOURNAL	1 (bases 1 to 27166)		
REFERENCE	Bowman,S.		
AUTHORS	2 (bases 1 to 27166)		
JOURNAL	Barrell,B.G., Rajandream,M.A. and Walsh,S.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUN-1995) Saccharomyces cerevisiae Chromosome XVI		
	sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge		
	CB10 1RQ E-mail: barrell@sanger.ac.uk		
COMMENT	Notes: All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis. Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS. Cosmid 9531 is overlapped at the 5' end by cosmid 9723, emb1 entry SC9723, accession no. Z48951, and at the 3' end by cosmid 9367, emb1 entry SC9367, accession no. Z49274.		
FEATURES	Location/Qualifiers		
source	1..27166 /organism="Saccharomyces cerevisiae" /mol_type="genomic DNA" /strain="AB972" /db_xref="taxon:4932" /chromosome="XVI" /map="16R" /clone="cosmid 9531" c1..1727 /note="YP9531.01, unknown, len: 574, CAI: 0.14, overlaps and extends cosmid 9723 orf YP9723.08 which has similarity at N-terminus to transcription factors" /codon_start=3 /product="unknown" /protein_id="CAA90152.1" /db_xref="GI:887585" /db_xref="GOA:Q12753" /db_xref="UniProt/Swiss-Prot:Q12753" /translation="SISRQPMFVERLVSTSLDSNLSGHGALSDTSSILRSTFLDSE PEVGKISDYHHPVSLAGISLSLQSSQSDQNFPSIQSPPLSGMSFNLGTINETNQN HSNHQSKSGNNWDSVSLPAKADSLNNMDKNSVGLDLGHGSKRISPSKNSRVGE VSVPLEEIPSDIDGVRVTDKSLVYDPDESIERNFSTATAATGESKEDINDNC NRNSKYSKTNMNGMNNNSNNNSGNDKNNNSRQEHQGLDFMFDITDSS ISTLSRANLLQEKIGSQNSVKQENSKNPKLRHLTSRSRSTFIHHPANBYLNTFG NSHNDIGKGVELVLTSPSFMIDPEKERETERSPSNYITDRPTRKRSSSIDVNRH YPPMADTVATSPGALNNAASNLDDQLSLTSLNSQPSIIAMMDPSNLAEQSSIHS VPOSINSRPMPTGSRQDKNIHTKKEENPLNNIHLDSLENVDEMNMQFSPBLKSM NRPDARENSSSNFIIQNSMISTPSPGENDLPDTPSPSSIQTASPPSQLLTDQFAD LDNFMSSL"		
CDS	1..103 /note="overlap with cosmid 9723, positions 17558 to 17660; EMBL:SC9723, Z48951" 3177..3983 /note="YP9531.02 unknown, len: , CAI: , possible zinc finger protein" /codon_start=1 /product="unknown" /protein_id="CAA90153.1" /db_xref="GI:887586" /db_xref="GOA:Q12286" /db_xref="SGD:S000006213" /db_xref="UniProt/TREMBL:Q12286" /translation="MKPNRRTCDVITNRDESLPALLLPALNSYTCDDSLKKGQISSNG RYQPFQFSDCSLLPKRLNIQAGQGMVSSIQCADHSYNNQKSEKTKPKLGGCPTE YTEYKYTSGETTDSAVVSIATNRLKRRQDQPCSDSCRKIKCNATITIIPLQD RNLSSISSNLHYLISQDDINQPRMKPFRKLPDVWGTVEVIKHLDKIVLPKACTSCSR		
misc_feature			
CDS			
misc_feature			

	RNOKNGKCLFSRGFTKSDMNVFPKINSKLDKDKSIFEMTVDDYVAAQFOTL"		
	complement(4210..7821)		
	/gene="RPA2"		
	complement(4210..7821)		
	/gene="RPA2"		
	/note="[9531.03c, RPA2 gene (RPA135, SRP3, RRN2)		
	SW.RPA2 YEAST P22138 DNA-directed RNA polymerase I 135 KD		
	polypeptide (EC 2.7.7.6) (A135)(RNA polymerase I subunit		
	2)"		
	/codon_start=1		
	/product="Rpa2p"		
	/protein_id="CAA90154.1"		
	/db_xref="GI:887587"		
	/db_xref="GOA:P22138"		
	/db_xref="InterPro:IPR007120"		
	/db_xref="InterPro:IPR007121"		
	/db_xref="InterPro:IPR007641"		
	/db_xref="InterPro:IPR007642"		
	/db_xref="InterPro:IPR007644"		
	/db_xref="InterPro:IPR007645"		
	/db_xref="InterPro:IPR007647"		
	/db_xref="InterPro:IPR009674"		
	/db_xref="SGD:S00006214"		
	/db_xref="UniProt/Swiss-Prot:P22138"		
	/translation="MSKVIKPPGQARTADFTLRESRFINPPKDKSAFFLLQRAVQP HIGSFNALTGPDGGLNLGVKDIIGERKIFDGKPLNSEDEISNGYLGNKLSVVEVQ SIAPKMSNDGVSASVAKVYPSRSEKRLTSYRGKLLKLKWSVNGNEELFEVRDCGG LPVNLQSNRCHLNKMSPYELVQHKESDEIGGYFIVNGIEKLIIMLIYQRNHPMAII RPSPANRGAASYHYGIIQIRSVDPQTSQTNVLYLNDGOVTFRFSWRKNEVLYPVMI LKALCHTSREIPDGIIGNDVKDSFLTRLELLRGLPKKRYPHLQNRTOVLQYLGDKF RVFQASPDQSDLEVGQEVLDRIVLVHLGKDSQSKRMLLFMIRKLSLVAGSCSPD NPDATQHEVLGGLYMLKEKIDISYLNIIAQVRMDINRGMAINFKDKRYMSRVL MRVNETIGSKMQYFLSTNLVSGSLDQQVSYVVAEKINFRFTSHFRVHRGSF FAQLKTTVTRKLLPESWGFCLPVTDPGSPCGLLNHFAHKCRISTQQSDVRIPSLY SLGVAPASHTFAAGSLCCQIDGKIIGWVSHEQKIIADTLRYKVRGKTPGLPIDL EIGVVPSTRGQVPLYLFGCHSMLRPVRLPLDKEDIVGPFQVYMNIAVTPEIO NNVTHVFPTNLSILANLTSPSDFNQSPRNNYQCOMGQKTWGTGVALCHSDNK LYRLQGTQPLVKANLYDDYDMDNFPNGFNAAVAVISYGYDDMDAMIINKSADERG GYGMYTEKVDLALNRNRGDIPTQHFEGNDKPWKLEKLDGELGYIETTYVEEG PICAYFDDTLNKTIKTYHSSEPAYIEVNLIGDESINKFQELQVTSIKYRIRRTPOIG DKFSRRHGQGVCSRKWPTIDMPSFTGIQDPIIINHAPFSMTIGMFVSLAGKAG ALHGIAODSTPMIENEDDTPADYFEGOLAKAGYVNHGNEPMYSATGELRADIYVG VYQRLRMVNDKFEQVRSSTGPNVSLTWQPVKGRKHGIRVGEMERDALIGHGTSFL QDRLNSDSTYQASVCRECGSILITQQSPRIGSISTVCCRRCSMRFPDACKLLTKE DGEKIFDSTQIWDGQGNKFPVGNQNTTTVAIPFVLKYLDSELSAMGIRLRYNVEPK" 8687..8759 /product="tRNA-Lys" /note="tRNA-Lys gene, anticodon ctt, len: 73" complement(9687..10667) /note="YP9531.04c, similar to SW:GDC_HUMAN P16260 Grave's disease carrier protein (31.0% identity in 300 aa overlap); contains 2 x PS00215 Mitochondrial energy transfer proteins signature" /codon_start=1 /product="unknown" /protein_id="CAA90155.1" /db_xref="GI:887588" /db_xref="GOA:Q12251" /db_xref="InterPro:IPR001993" /db_xref="InterPro:IPR002067" /db_xref="InterPro:IPR002113" /db_xref="InterPro:IPR002115" /db_xref="SGD:S000006215" /db_xref="UniProt/TREMBL:Q12251" /translation="MAEVLTVLEQPNISKDFLKQDSNIAFLAGGVAGASVTRTVSPFFE RVKILLQVQSSTTSYNGRIFSSIRQVYHEGTGKLFGRGNLNCIRIPFSAVQFVYVE ACKKLFHVNGNNGEQLTQRLFSGALCGSCSVATYPLDLIKTRLSIOTANVSSL NRKAKSISRPDGPQWLLSETYRLEGGLRGLYRGVWPSLGVVPYVALNFAVEQGLRE FGVNSDAQSPKSWKSNLYKLTIGALSQGVQAQITTFDILRRRFQVLNAGNELGRYT SVMDALVTIGRAEGVSGYIGLAANLKFVWPSTAVSLVIEYVCDVSRNW" complement(10218..10244) /note="PS00215 Mitochondrial energy transfer proteins signature" complement(10518..10544)		


```
/note="PS00215 Mitochondrial energy transfer proteins
signature"
10934..11188
/note="YP9531.05, unknown, len: 84, CAI: 0.12"
/codon_start=1
/product="unknown"
/protein_id="CAA90156.1"
/db_xref="GI:887589"
/db_xref="SGD:S000006216"
/db_xref="UniProt/TREMBL:Q04203"
/translation="MELSOYLNYAFSLAYIIIIHLLCLSYIYEIIHKHKNVPRPSKL
EALPYKGNKNEGDRLSIFPLLLKSVKHSIRVT"
complement(11257..12210)
/note="YP9531.06c, unknown, len: 317, CAI: 0.13, putative
zinc finger protein, contains PS00028 Zinc finger, C2H2
type, domain"
/codon_start=1
/product="unknown"
/protein_id="CAA90157.1"
/db_xref="GI:887590"
/db_xref="GOA:Q12145"
/db_xref="InterPro:IPR007087"
/db_xref="SGD:S000006217"
/db_xref="UniProt/TREMBL:Q12145"
/translation="MNFSLSKSQSEKQSYDKSRSPNIGMCTVMYKSNLPLQVSAVD
QLSTGKGFQYGVNKEIPITASEPNTALSMHFNALSLCHDRPKSVPRGYGLPTMOE
PVYKPLPYDKERITIASERPNFALSMHFNALSLCHDRPKSVPRGYGLPTMOE
ASNGRTKSEPGAVLNGSATPSDKWTDTRISASHKLRQCPVKCKISRPSTLKHLYIH
TGDTPFKTWBECTKSFNVKSNLRLKSHERRKNKVLNTT"
complement(11293..11361)
/note="PS00028 Zinc finger, C2H2 type, domain"
complement(11383..11445)
/note="PS00028 Zinc finger, C2H2 type, domain"
complement(13814..14143)
/note="YP9531.07c unknown, len: 109, CAI: 0.08"
/codon_start=1
/product="unknown"
/protein_id="CAA90158.1"
/db_xref="GI:887591"
/db_xref="SGD:S000006218"
/translation="MSCIFSDLGVYSCABSRITNLVLCIICIREEKAAPVVKDKF
LFFIFLSSKENLFCFISRSRSEVSKLRKNFICPHAVIGFOGAKAQLGRFPLN
GLRLQH"
complement(16908..17651)
/note="YP9531.08c, unknown, len: 247, CAI: 0.11, putative
zinc finger protein, contains 2 x PS00028 Zinc finger,
Alignment Scores:
Pred. No.: 1e-58 Length: 27166
Score: 1833.00 Matches: 393
Percent Similarity: 59.36% Conservative: 178
Best Local Similarity: 40.85% Mismatches: 281
Query Match: 39.77% Indels: 110
DB: 15 Gaps: 14
US-10-768-511-8 (1-901) x SC9531X (1-27166)
QY 2 GluAsnAspAlaLeuAspIleGlyAlaValSerProTyrProSerGlnSerGlu 21
Db 23405 GAGGATAATAACTCC-----AGCTCCCGCTTGTGCTTAATCCTGAT 23446
QY 22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaValSerPro 41
Db 23447 TCTGTTCCTCA-----CCACAGCTTTCTTCCCGAGCTCTATTTTATAGCTCCTCTCA 23497
QY 42 ValAlaGly----- 44
Db 23498 TCACAGGTGATATCTATGTGTCGCAACATAGCCAGAACTTAAGTCAGGAGGGGAAAC 23557
QY 45 ---ArgArgAlaValArgGlnThrPro-----ThrSerAlaValArgArgGly 60
Db 23558 ATCAGAGCTGCTATAGTCTTCTTCCCAATAATTTTCCATCTTCTTCCCAAGACAAAT 23617
```

```
QY 61 ArgGluThrAspSerAlaArgArgSerArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db 23618 TCCGATGTTTCCCAATCTCAAGGCAGACAGGCAGAAATCGTCTTCTTCGCGAGCGCTCT 23677
QY 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProGlyAlaThrPro 100
Db 23678 GGAAGGTCTAGATATCATCTGATCTGAGAGTGTAGAGCACTGCTTCTTCTTCC----- 23734
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db 23734 ----- 23734
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
Db 23735 -----TCTTCTTTAGCGGTATGCTCAAAACCGTGTACACATCGCAAGA 23779
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db 23780 AATGATATTCATACATCTGATTTTCTCTCCAGAAGAATTTGTGGATTTTGATACATGA 23839
QY 161 -----ProSerGluSerAlaGlu 166
Db 23840 TCCGGCGTGAATCTTTGGATACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 23899
QY 167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGlyThrAsnValAsnIlePro 186
Db 23900 CCC-----TTGAGATAAATTTGGGTACCAACCGTGTAGTATCCAG 23938
QY 187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg----- 200
Db 23939 GAATGTACAACTAATTTTTCGTAATTTTTCGTAATTTTTCGTAATTTTTCGTAATTTTTCG 23998
QY 201 -----SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
Db 23999 TTGGATGAAGGAGGAATTCATTACATATCTACCGACGAGAACTATCTATATCAAG 24058
QY 213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
Db 24059 CAGCTTAATGAATGAGAGAACTTGTGTACCTCTAATTTTAACTTGTGATGTAGAAACTTA 24118
QY 233 -----TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
Db 24119 CTTGCTTACAACAAACAGAGAACTTATACCACTCACTAATTAATTAATTAATTAATTAAT 24178
QY 251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu----- 265
Db 24179 ATTTCTATCATGATCAAACTATCAAGGACTGTATGTTTCTTATATAGTAGCAACAACT 24238
QY 266 -----LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
Db 24239 TTGGATTACGATTTAGATGAATAGAGACCAAGTTTATAAGAAAGGCTTACAATGTA 24298
QY 281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db 24299 GGTTCCTGTAAGGTATGCGCAATTTGAATCCAAATGATATGATAATTAATTAATTAATTAAT 24358
QY 301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db 24359 AAGGGTCTTGTCTGAGATCAACTCCGGTTATCCCTGATATGAAGTAGGCTGTTTTCAAA 24418
QY 321 CysLeuValCysGlyHisSerProLeuValThrValValLysGlyArgValGluGlu 340
Db 24419 TGCAAGCTCTGCGATCATACA---ATGGCAGTGGNAATGATAGAGGATTAACAAGAG 24475
QY 341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db 24476 CCGCTAGGTGTGAACGATTTGATTTGATTAATGAACCAATTTCCATGTCTGATTCACAC 24535
QY 361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db 24536 AGGTGTTTCATTCGACATAAACAAAGTCATTAAGTTACAGGAAACTCCAGACTTTGTGCTT 24595
```

QY 381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db 24596 GATGACAAACGCCCTCATCTATTCATTTAGTGTGTTACGATGAATTAGTGATCTTTGC 24655
QY 401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db 24656 AGGCGGGGATCGTATTGAGTGACTGCGCAGTTTCAGGTCCATCCCATAGAGCTAAT 24715
QY 421 ProAsnGlnArgThrLeuArgAlaLeuTyrIleThrValAspCysValHisValLys 440
Db 24716 TCCAGGCAACGGGTACTAAAGTCGTTGTATATAAACATACGTCGATGTCGTCACCGTTAA 24775
QY 441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db 24776 AAAGTTTCAGATAAGAGGTAGACGTCGATCTTCTACTATTATGAACAAGNAATTAATGCAG 24835
QY 461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db 24836 AACAGGTAGATCATACGAGTCGAGGTGCAAGNAGTAAGACAAATTAATGATCAGGATTAGCA 24895
QY 481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db 24896 AAAATTCGAGGTTCGCGCAAGAGAGATTATATACGTTTATAGCCGCTCTATTGCC 24955
QY 501 ProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db 24956 CCAAGTATTATACGAGCTAGAGATGTCGAAGAGGTATATTACTTTCAGCTATTTCGGCGC 25015
QY 521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db 25016 ACGATAAAACCTTTACCAGGGTGGCGGTATAGAGGTGACATAAATTTTACTTTGT 25075
QY 541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db 25076 GGGGATCCTTCTACTTCCAAATCGCAAAATTTTGCAATACGTTTCAAAATTAATCTCCTCGT 25135
QY 561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db 25136 GGTGTGTATACTTCGGGTAAAGGTTTCATCTGCGGTGTTTAACTGCTTATATACAAGG 25195
QY 581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db 25196 GATGTCGACAAACAACTGTTTGGAAAGTGTGCTAGTATTGCTGTGAGAGGT 25255
QY 601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaAaArgSerMetLeuHisGlu 620
Db 25256 GTTTCGTGTATTGACGAGTTTGATAAATAGTGTATCTACAAGTCCGCTTCGACGAA 25315
QY 621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleAlaSerLeuAsnAla 640
Db 25316 GTCATGGAACAGCAGACTATTTCATTCGCAAAAGCGGGAATTTATCACACACTTAATGCC 25375
QY 641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db 25376 AGAAGTTCTATTTCGCCAGTGTCTAACCCAAATGTTTCAGCTACACCCCAATTTGGCT 25435
QY 661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db 25436 GTGACTGAAATATTGATCTACCGCCCTTACTTTCGAGATTTCGATCTGCTATCTT 25495
QY 681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db 25496 GTCTCTGTATAGTTGTATGAGAAAATGACAGAACTAGCCAAACACTTAACAAATCTT 25555
QY 701 HisTyrGluAsn-----TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu 718
Db 25556 TACCTGGAGATAAGCCCGAAACATATTCTCAAGCAGCTGCTACCAAGTTGAATTTTA 25615
QY 719 ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
Db 25616 ACGATGTATATTAGTTATGCAAGGAGCAGCATACACCCCAATAATTAATCACCGAGGCCGCTAAG 25675
QY 739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758

Db 25676 ACTGAGCTTGTTCTGCTGCTTATAGTAGAAGATGGAAGATGGGTGACGATCCGAT 25735
QY 759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
Db 25736 GAGAAGAGATCACAGCTACCCACACAGCAACTTCAAAAGTATGATTGCTTTGGTGGCGG 25795
QY 779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
Db 25796 CACGCCAAAATGAAATGAAAAACGTCGTAGAGCTGGAGGATGTTCAAGAAGCCGTTAGA 25855
QY 799 LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet 818
Db 25856 TTAATTAGATCAGCCATAAAGATTATGCAACACCCCTAAAACCGTAAATTCGACATG 25915
QY 819 AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAla 838
Db 25916 AATTAGTTCAAACAGGTAAATCAGTTATTACAGAGAAAACCTACAGGAGGATTTGTCAAGG 25975
QY 839 AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThr 858
Db 25976 GAAATATTGAATGTTTTCGAAGGATCAG-----GCATCAGACTCAATGTCTATTC 26023
QY 859 SerGlnLeuLeuGluAspIleArgSerGlnSerValAspValSerLeuGlnAspIle 878
Db 26024 AATGAGCTTATAAAACAAATCAATGAACACTCTCAAGATAGATGTGACTTCTGTATATC 26083
QY 879 LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
Db 26084 CAGGAAGCCTTGTCAAGATTGCAACAGAGACAAGGTCATGTCCTTGGCGAGGGTCTA 26143
QY 899 LysArg 900
Db 26144 AGGAGA 26149
RESULT 12
SCCHRXVI
LOCUS 165536 bp DNA linear PLN 18-APR-2005
DEFINITION S.cerevisiae chromosome XVI 165536 bp sequence, cen rightwards.
ACCESSION Z71255
VERSION Z71255.1 GI:1279666
KEYWORDS AR07; ATPase; beta-transducin; CCL1; CDC21 homologue; CDE1; CDE2; CDE3; centromere; CHL1; chorismate mutase; citrate synthase; coatomer zeta subunit; cyclin; delta element; DnaJ; DSS4; electron transport flavoprotein; eukaryotic initiation factor 5; F1Fo-ATP synthase g subunit; gamma adaptin; GNL1; glutamine synthetase; Grave's disease; guanine-releasing protein; HAL1; halotolerance; histidine tRNA synthetase; HTS1; isocitrate lyase; LPT1; LTR; MAK3; MFS1; Mitochondrial carrier protein; N-acetyltransferase; NADPH-cytochrome P450 reductase; negative regulator; NHP6A; nonhistone chromosomal protein 6A; OSD1; phenylalanyl-tRNA synthetase; phosphotyrosine protein phosphatase; polyposis; proteinase P1; protein kinase; protein phosphatase; regulatory protein; ribosomal protein L37a; RNA polymerase I; ROX1; RP2; SEC8; SMK1; spermidine synthase; TIF5; TKL1; transfer RNA-Gly; transfer RNA-Lys; transfer RNA-Phe; transfer RNA-Ser; transketolase; tropomyosin; ubiquitin-activating enzyme; vacuolar H(+)-ATPase 54 kDa subunit; VMA13; YMC1; YME1; YTA11; zinc finger protein.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 165536)
AUTHORS Badcock,K., Bowman,S., Churcher,C.M., Pearson,D., Rajandream,M.A., Walsh,S.V. and Bartell,B.G.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) Saccharomyces cerevisiae chromosome IV
CB10 1RQ. E-mail: barrell@sanger.ac.uk
COMMENT SC8132X 248483 1 to 25540
SC9723 248951 100 to 17660
SC9531X 249319 104 to 27166

SC9367 249274 10 to 42190
SC9385 268111 790 to 10339 (PCR product)
SC9499X 249219 134 to 43776

The most significant matches using FASTA -o to a non-redundant database compiled from Swissprot 32 and PIR 46 and Wormpep9 are included.

All CDS over 100 codons have been analysed. CDS that are largely or completely overlapped by a larger CDS have been omitted from this analysis.

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons/amino-acids is given for each CDS as is the calculated codon adaptation index (CAI).

tRNA genes and LTRs are also included.

Notes:

This sequence has been compiled from the following submissions of individual contigs.

Location/Qualifiers

1. .165536

/organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XVI"

<1. .3032

/genes="YPL012W"

<1. .3032

/genes="YPL012W"

/note="YPL012W, unknown, incomplete, len: 1009, CAI: 0.19"

/codon_start=3

/product="unknown"

/protein_id="CAA95029.1"

/db_xref="GI:1314064"

/db_xref="GOA:Q7LH12"

/db_xref="InterPro:IPR000357"

/db_xref="UniProt/TREMBL:Q7LH12"

/translation="SNLSNKLQAQTKEDINASVNRSLRLTSVVSQTQSPSSQIEP

LCVLLVQKTSQSVIVASFCESFMFKTAETITSSGAENKYLRLVDTIPALPKS

NVDLTLLKSWIAVIMSTYATHPLKALRKIPGVHIMCTVLASETPEVYQAASQC

LILSLESVKDILLTPSVDEKVNDELLISQIAKTFIDFLSIRYSHCSREILKIL

VAAFNFRSRPHFLUKLIVTVRWNEEQFMDLNEIELVIGASISANGPEMILAE

APLNLDPSRPGRAWLLPLINDYTKNANLATFQNELAYIKSFQKFDKVPESIQ

LRVQIVQIWSLTPFCELPMDRESPTDEPASELSLLYSEVLRITICHALKV

ASNSVYQIESSHNVLLQRPPISEAQNIIEYLTKSTNLLAVLFNVYTTQTPNARS

YILETDQYLTKSKEDLKTENNVCGLKSNWRESSGNVKEKKKPPOLATLDDLI

ICMTIYVSSYSALFSMESLTVNSADALIQRAYRIIFKLSELKSGSTAVAQFISDI

ENWVDSASVQTSAAARLAIKTIIVELLPLDHLDFIVRTVAEVLSTKDVNKEKRE

TAFDTLCGRKMEPNIGIKLQIPGYDPTPDQSSISEPFKIIISAGLIGESQHW

SSSITGACLVLFKNEVLSGILMDIYDTIELYLSNREIVKSAIGFTKVCYLGLPE

ELMRPVACPLLLKLLWSHEHTGHFRAKVKHIIERLIRRGVDYIEANPEEDRLLT

NIRKMNRRKRDDEVTTCVSQVAATKGSRFMSAPDEAVYQSDNEEDNGSDOEENAG

GRMNGAKQFIVESGNPLDLSQTLAHISSTRPKFKNKQNRARPNDDAFNPDSEG

DLVVGQKPESTNVDDPLSAVTGGINAYLEAVKSGVPGVQRNKLKFRKNGKSDSEFGD

DDGEXKDSRLMRGRVQGNKIGIKHKKNGKPKFKSRKL"

complement (3265. .4326)

/genes="YPL011C"

complement (3265. .4326)

/genes="YPL011C, unknown, len: 353, CAI: 0.12"

/codon_start=1

/product="unknown"

/protein_id="CAA95030.1"

/db_xref="GI:1314065"

/db_xref="GOA:Q12297"

/db_xref="InterPro:IPR006565"

/db_xref="InterPro:IPR009072"

/db_xref="SGD:S000005932"

/db_xref="UniProt/Swiss-Prot:Q12297"

/translations="MTNNDFFALLRISILQLLKAQGFDRAPSLVDVMTDLAYAKFL

SLIASVSSIAQARCDQDDTIALQDITALENIGIVKPTNVLVDVENSELSSSGME

KFKWDCTYDARITATLPTVLESEKESDPLSAIPYLNQLNKKGAQKQLEIT

KNRKTLIEDLINNGLDWIKLVIARQINMIERASKKESQNVLPALPIAGYKSSIL

SRHHHTTITNEDRMPSAMTPRDEDALEIQENPFVTSKLPIMEKRNLENITLSPEDE

gene

CDS

ELBSLGEVGPQKQSQENNNERSFKENNKSLTESPHGDDRDISMFPQDSNVDTKWAEQ
EDMDSTFORRTSLDYGYP"

4641. .5210

/gene="YPL010W"

4641. .5210

/gene="YPL010W"

/note="YPL010W, unknown, len: 189, CAI: 0.19, 39.7%

identity in 156 aa overlap to coatomer zeta subunit

COPZ_BOVIN P35604"

/codon_start=1

/product="unknown"

/protein_id="CAA95031.1"

/db_xref="GI:1314066"

/db_xref="GOA:P53600"

/db_xref="InterPro:IPR000804"

/db_xref="InterPro:IPR011012"

/db_xref="SGD:S000005931"

/db_xref="UniProt/Swiss-Prot:P53600"

/translation="MSSLSLTVQAVLLDQGERIYAKYQPPHRSDGHQLLFNSV

KKQFEKQLYRTHKQDSEILIFEDHLVLYKEYIDITTYLVASLEENIVLQQGSA

IRGALDILNSGMKKNIOENYDMVLLAIDETIDNGVILETDSNTIASRVSKPPTNEP

QNALDLDKGFLGAWGFAKSKFOERLQQL"

complement (5443. .8559)

/gene="YPL009C"

complement (5443. .8559)

/gene="YPL009C, unknown, len: 1038, CAI: 0.19, contains

glutamic acid rich region"

/codon_start=1

/product="unknown"

/protein_id="CAA95032.1"

/db_xref="GI:1314067"

/db_xref="GOA:Q12532"

/db_xref="InterPro:IPR008532"

/db_xref="SGD:S000005930"

/db_xref="UniProt/TREMBL:Q12532"

/translation="MKRISALDLLLLELQDLLEGYRLSNIYNTADSKQFLLPKN

KPDSLNVVDCGLRIYLFTEFRPIPTSPGFVKLRKHLKAKRLTALKQVDDRLIV

LQFADHPVLVLEFFSAGNVILLDENRRIMALQRVVLEHENVKQIYEMFDSLSFTTN

NSSADETEKNEKAEVTSVLNEMIKAVQAKYESDITVIKQNIQKQKAKKKVVP

SHKLLLSKPHLSDDLKSNLVNIDPSESCLNLEETDSLAELNSTQVLEYNOLL

TTTDKGYILAKRNEYISEKOTADLEFIYDFPKPYINGDTSOCCIEVEGPTN

RTLDPIETESSKALRIQOESQAQKIDARAEKRIQALLDVQELNERKHLI

IKNALPIEBVLAVQGLIDQDMWNTIEKLIKEQKGNRIQALLNPLNKQKLSV

SEKELANTSDENESGNTTSSSDSDSDMBSSKSTKSKRKSNEKINVTI

DGLSAYANATEVNIKKTSAOKKVKNGVKAMKNIKVIDQQLKKLQDSHVLK

KIRTPFFKYSWFFISSEGLFVMGKSPAETQIYKYIEDDDIYMSNFSNHWIKN

PKTEVPPNLMQAGILCMSSSEAKSKISSPFWCFAKNVSKFGSDSDSILPEGAFR

LKNENDQNLPPAQLVMGFLWKVKTSGNEDGDDDEEEEEEEEEEEEEEEEEEE

EEKEEEEEEEQQDDEDSNEVNGLEKGDSDNDSTKNSFEHNLKEDIKKTCTISS

DTSDSGNAKAKNDNSTORILDEPGVPSLIENINSNVKRGKGLKIQKQYADQDE

TERLLRLEALGTLKGLKQKQKKEIKMKREVRDRKREKORRLOALKFTKKEKAR

YNDKHKLSLKSLDKGVDDDIIPVFAFPALLKYKYKVKIQPSAKKTKTLEILH

YFKPSLDGSDSTDNENDWPQHEMIKGLKEQDLVLLLCVDLKVITAGQSTKNGNS

SKKGKKKR"

9008. .11593

/gene="CHL1"

9008. .11593

/gene="CHL1"

/note="YPL008W, CHL1 gene, len: 861 CAI: 0.11

SW:CHL1 YEAST P22516, CHL1 protein, contains P50690

DEAH-box subfamily ATP-dependent helicases signature"

/codon_start=1

/product="Chl1p"

/protein_id="CAA95033.1"

/db_xref="GI:1314068"

/db_xref="GOA:P22516"

/db_xref="InterPro:IPR001945"

/db_xref="InterPro:IPR002464"

/db_xref="InterPro:IPR006554"

/db_xref="InterPro:IPR006555"

/db_xref="InterPro:IPR010614"

/db_xref="SGD:S000005929"

gene

CDS

```

/db_xref="UniProt/Swiss-Prot:P22516"
/translation="MDKKEYSETFVHPYKPYDIOVLQVLMETVVRVLSGKXIALESPT
GTGKLSLICATWMLRMNKADIFRMEETNIKTNEHDSSENISDDPEPDWIDTYRKSVL
QKVDLLNDIEKHLNEINTISCKQLKTMCDLGDGHEGRYKSVDLPKRGARHLDVSL
BEQDFIPRPYEDSNNDTSKSTGRISDKDYKLSNELNSQIITLLDKDKGVSRDPN
NGRFDVTPNQPKVIYYASRYSQLQFSLRPLFPSPFSFRDKVPDKVKYLPPLASK
KQLCINPKVMKWKLEADACADILRKEGCIQYONTNEWRHCPDTLALRDMIFSEI
IDIEDLPKSLGICPVYASREALPIAEVUTLPIVYLLSSESTRSSLOINLENSIVII
DEAHNLITINSIYSQISLEDLKNCHKGIYTFNFKFSRLPNPGRNVNLLKLSLMT
LQFIVNPKIKGQIDPDMFTGSNIDTLMHLRLRYIKVSKIAKIDITNQALKEE
ESSKNENPIKTHKSSQSPQLLFVQSQFLYCLNTLNTSEGOFFFKFNYSIKMLLEPS
KPFESILNKAQCVLAGLMEFMSFSLNLLPEVSEDIITLSCNVHPIKENLQYIIT
NQPELEPTFEKMSPLVNNHLFQFVLSKAVPKKGGIVAFPPSYQYLAHVIOCWKO
NDVPATLNVKRIYEAQXGDDILSGYSDSVAEGRGSLLLAIVGKLSGINFDDLC
RNVVMVGI.PPFPNI.FSGELIVRKHLAAKIMSGGTEEASRAATKEPMENICMKAVNQ5
VGAIRHANDYANIYLLDVRNPNFRKRLSGRWQDSINSEHTHQVLSSTRKPFMSR
SLNSR"
misc_feature
10169..10198

```

Alignment Scores:

```

Pred. No.: 6, 01e-58 Length: 165536
Score: 1833.00 Matches: 393
Percent Similarity: 59.36% Conservative: 178
Best Local Similarity: 40.85% Mismatches: 281
Query Match: 39.77% Indels: 110
DB: 15 Gaps: 14

```

US-10-768-511-8 (1-901) x SCCHRVI (1-165536)

```

QY 2 GluAsnAspAlaLeuAspIleGlyAlaValSerSerProTyPrSerGlnSerGlu 21
DB 66403 GAGGATAATACTCC-----AGCTCCCGGTGTGCTTAATCCTGAT 66444
QY 22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaSerPro 41
DB 66445 TCTGTTCCA-----CCACAGCTTCTCCCGAGCTCTATTATTATAGCTCTCTTCA 66495
QY 42 ValAlaGly-----44
DB 66496 TCACAAAGGTGATPATCTATGGTCGCAACAATAGCCAGAACTTAAGTCAGGAGAGGAAAC 66555
QY 45 ---ArgArgAlaValArgGlnThrPro-----ThrSerAlaValArgArgGly 60
DB 66556 ATCAGAGCTGCTATAGGTCTTCTCCACTAAATTTTCCATCTTCTCCCAAGACAAAAT 66615
QY 61 ArgGluThrAspSerAlaArgArgArgSerArgSerLeuGlyAsnSerVal 80
DB 66616 TCCGATGTTTCCAAATCTCAAGGCACAGAGCGCAGAAATTCGTTCTTCTCGCAGCGCTCT 66675
QY 81 TyrSerSerProTyAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
DB 66676 GGAAGGTCTAGATATCAATCTGATCTGAGAAGTGTAGAGCACTGCCACTCTTCTCC--- 66732
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
DB 66732 -----66732
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
DB 66733 -----TCTTCTTTAGCCGTAATAGTGTCAAAACCGTGTACACATGCGAAGA 66777
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
DB 66778 AATGATATTCATACATCTGATTTATCTCTCCAAAGAAGAAATTTGTGGATTTTGTACTACTAGA 66837
QY 161 -----ProSerGluSerAlaGlu 166
DB 66838 TCCGCGGTGAATACCTTTGGATCTTCTTCTTCTTCTGCTCTCTCCATCGGAAGCCAGTGA 66897
QY 167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValThrGlyThrAsnValAsnIlePro 186
DB 66898 CCC-----TTGAGAATAAATTTGGGTGTACCAACGTCATGATATCCAG 66936

```

```

QY 187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg-----200
DB 66937 GAATGTACAACATAATTTTCGTAATTTTGTATGCTCTTTCAAGTATAAATTCGTTAAATA 66996
QY 201 -----SerSerAlaHisLeuAsnSerLysTyrIleGln 212
DB 66997 TTGGATGAAGGGAGGAATTCATTAAACAATACTACCGACGAAGAACTATATCTATATCAAG 67056
QY 213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
DB 67057 CAGCTTAATGAATGAGAGAACTTGGTACCTCTAATTTAAATTTGATGCTAGAACACTTA 67116
QY 233 -----TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
DB 67117 CTTGCTTACAAGCAACAGAGACTTATACCATCAACTATTAAATTAATTCCTCAAGAGTG 67176
QY 251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu-----265
DB 67177 ATTTCTATCATGATCAAACTATCAAGGACTGTATGGTTCTCTTAATAGTAGACAACAAT 67236
QY 266 -----LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
DB 67237 TTGGATTACGATTTAGATGAATAATAGACCAAGTTTATTAAGTAAAGCCCTTACAAATGTA 67296
QY 281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
DB 67297 GGTCTTGTAAAGGTATGCCGAATTTGAATCAATGATATTGATTAATTTGATAAACTTA 67356
QY 301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
DB 67357 AAGGCTCTGCTCGATCAACTCCGCTTATCCCTGATATGAAGTAGGAGCGTTTTTCAAA 67416
QY 321 CysLeuValCysGlyHisSerProLeuValThrValValLysGlyArgValGluGlu 340
DB 67417 TGCACGCTCTCGCATCATACA---ATGCGACGTGGAATTTGATAGAGGAGTTATACAAGAG 67473
QY 341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
DB 67474 CCCGTAGTGTGAACGTTATGATTTGAATGAACCCCAATTCATGCTACTGATTCACAAAC 67533
QY 361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
DB 67534 AGGTGTTTCATTCGCAGATAAACAAGTCATTAAGTTTACAGGAACCTCCAGACTTTGCTGCT 67593
QY 381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
DB 67594 GATGGACAAACGCTCACTCTATCTCATTTATGTGTTTACGATGAATTAATGATGATTTCTG 67653
QY 401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
DB 67654 AGGGCGGCGATCGTATTGAGTGACTGCGCAGTTCAGGTCCATCCCATTTAGAGCTAAT 67713
QY 421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
DB 67714 TCCAGGCAACGCTACTAAAGTCGTTGTATAAACATACGTCGATGTGTCACGCTTAA 67773
QY 441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
DB 67774 AAAAGTTTCAGATAAGAGGTTAGACGTCGATCTCTACTATTGAAACAAGAAATTAATGCGAG 67833
QY 461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
DB 67834 AACAGGTAGATCATACGAGGTGCGAAGTAAGACAAATTTACTGNTCAGGATTTAGCA 67893
QY 481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
DB 67894 AAAATTCGAGGTTGCGCAAGAGAAAGATTTATACAGTTTATTAGCCGCTCTATTGCC 67953
QY 501 ProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
DB 67954 CCAAGTATTTACGAGCTAGAGATGTCAAGAGGGGTATATTACTTTCAGCTATTTCGCGCG 68013
QY 521 LysAlaLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540

```



```
Db 1918 GCAGAGGCTGGGATCATCTGTGCTCAATGCGCGCAGCTCTGTCTCTGCGCAGCAGCAAT 1977
Qy ProSerGlySerArgTyrAspAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
Db 1978 CCTATTGAGTCTCAGTGGAACTCTAAATAACAACCAATGAAATATCAATCCGCGAC 2037
Qy ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
Db 2038 ACATTGTTGTCAGGTTTGATCTCATTTTCTCATGCTAGACCTCAGGATGAGGCATAT 2097
Qy AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
Db 2098 GACCGCGCTAGCTCATCACCTGTTTCTATTGTAACCAAGTGAGGAGCAAGTGGAG 2157
Qy GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrAlaArgGlnHisVal 729
Db 2158 GAGGAGTTCCTGGACATGCCCTGCTGAAAGACTCATTTGCATATGCTCCATGATGACCATC 2217
Qy HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
Db 2218 ATGCCCCGACTGAGTGAGAGGCCAGCGCTCTCATTTGAGGCTTATGTAACATGAGG 2277
Qy GlnLysGlyAsnPheProGlySerSerLysValIleThrAlaThrProArgGlnLeu 769
Db 2278 AAG-----ATTGGGAGTAGCGCGGGGATGGTTCTGCTTACCCTCGACAGCTA 2325
Qy GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
Db 2326 GAGTCATTAAATTCGCTTAGCAGAGCCCATGCTAAAGTAAGATTTTCAACAAGTTGAA 2385
Qy LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
Db 2386 GCATTGATGTGAAGAGCAAAACCCCTCCACCGGAGGCTCTGAAGCAGTCTGCACCT 2445
Qy AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
Db 2446 GACCTCGTACTGCGATTGCGATATTTCTATTCTTACTACAGGATGATGCCACTTCT 2505
Qy ArgIleArgAlaAsnLeuAlaLeuArgGluLeuIleAlaAspLysIleSer 849
Db 2506 CGTAAACGGAAGAAAGATTAGTGAAGCATTTAGAGAAACTTATT-----TTATCT 2556
Qy ProGlySerSerGlyLysThrSerGlnLeuLeuGluAspIleArgSerGlnSer 869
Db 2557 AAGGGTAAACACACCGCTTAAAGTACCAACAGCTCTTTGAGGATATTCGGGACAGTCT 2616
Qy SerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
Db 2617 GACACAGCAATTACCAAGGACATGTTTGAAGAGCCCTCGAGCTTTGGCTGATGATGAT 2676
Qy PheLeuThrValHisGlyAspIleValLys 899
Db 2677 TTCCTAACAGTGACTGGGAAGACTGTCGCG 2706
```

RESULT 14

```
AK220442
LOCUS Mus musculus mRNA for mKIAA4003 protein. linear ROD 01-MAR-2005
DEFINITION AK220442
ACCESSION AK220442
VERSION AK220442.1 GI:60360103
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE
1 Okazaki, N., Kikuno, R.F., Ohara, R., Inamoto, S., Nagase, T., Ohara, O.
and Koga, H.
PREDICTION OF THE CODING SEQUENCES OF MOUSE HOMOLOGUES OF KIAA
Gene. The Complete Nucleotide Sequences of Mouse KIAA-homologous
cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
Randomly Sampled from Size-Fractionated Libraries.
```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Published Only in Database (2005)
2 (bases 1 to 3283)
Okazaki, N., F. Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
Direct Submission
Submitted (22-FEB-2005) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatari, Kisarazu, Chiba, 292-0818, Japan
(E-mail: mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)

FEATURES

Location/Qualifiers

1..3283

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="mfj10260"

/tissue type="fetal brain"

/note="Vector:modified pBC SK+"

<1..3283

/gene="mKIAA4003"

/gene="mKIAA4003"

<678..2711

/note="CDS is predicted by in silico analysis. Start codon

is not identified."

/codon_start=1

/evidence=not experimental

/product="mKIAA4003 protein"

/protein_id="BAD90271.1"

/db_xref="GI:60360104"

/translation="ILLNLCTNNLERLILQESHFLVNCHEHIKSFKNLYRQLISYP

QEVITPFMAVNEIFPDQYPCSHLEHQIQVRFNALTKTSMNLNPDIDQLITISGM

VIRTSLIPMEMAEAFPCQCVATHTVRIDRGRIAPSCVCHVTTHGSLNHRSF

SDNMKIQESPEMDPAGTPTIIVLFAHLDVQKQVQDGRVNVTVGIYRAVPIRVNPR

VSNKSVYKTHIDIVHRTDAKRLHGLDEAEOKLFSEKRVKLLKELSKPDIVRL

ASALAPSYEHEDIKKGLLQLFGCTKDFSHGTGKFAEINILICGDPGTSKQL

QIVNLVPRGVSITGKGSVAGLAIYVMDPTQQLVLTGALVSDNGICICIDEPDK

MNESTRVHVEVMEQTLISAKAGIICOLNATSVLAANPLESQWNPVKTTIENIQ

PHLLSRFLIFLMDPQDEAYDRLEHLVSLYQSEEVSEFLDMAVKDYTA

HSTIMPLRSEASQALIEAYVMNRKIGSRVMSAYPROLESILRLAEAAKVRFSNK

VEAIDVEAKRLHREALKOSATDPTGIVDISILTTGMSATSRKKEELAEALRLIL

SKGKTPALKYQQLFEDIRGQSDTAITKMFEBALALADDPLTGTGTVRLLL"

ORIGIN

Alignment Scores:

Pred. No.:	3 328-59	Length:	3283
Score:	1821.50	Matches:	403
Percent Similarity:	59.72%	Conservative:	133
Best Local Similarity:	43.22%	Mismatches:	265
Query Match:	39.52%	Indels:	110
DB:	9	Gaps:	17

US-10-768-511-8 (1-901) x AK220442 (1-3283)

Qy	14	SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro	33
Db	107	AGCAACCCAGCAGCACCATGTCGCCCGGCATCCACCCCG	145
Qy	34	SerPheAspAsnAlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThr	53
Db	146	-----AGCCGCGCAGCAGCAGCGCGGA---CGAGTCACCCCAACC	184
Qy	54	SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgSerArgSer	73
Db	185	CAGTCCCTTCGAAAGTGAGGAAAGAGGAGTCTCACCAATCGGAGACGTAGCGGAGAT	244
Qy	74	ArgSerLeuGly	77
Db	245	TCCTCCACCGGAGAGCTACTGCCAATGCCACCTCACCAGGACCGACCTGCAGAGCCCA	304
Qy	78	-----AsnSerValTyrSerSerPro	84
Db	305	CCTGCACAGAAATGCCCTGTTTCCAGCCCTCCTCAGATGCATTCCTTAGCTATTCCTTG	364
Qy	85	----TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla	103

Db 365 GACTTTGATGTTAGTTCACCATTTGACA-----TATGGC 397
Qy 104 ThrProValGlyThrProMetGlyThrPro----SerPheHisArgGlyThrProGlnTyr 122
Db 398 ACTCCAGCTCGCGAGTGGAGGAACCCCAAGAGTGGGTGGAGGACACCT---GTA 454
Qy 123 LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGln 142
Db 455 AGGCAGAGCCAGATCTGGCTCAGCAGAAAGGTTTGAGGTGGATCTGCAGTCT--- 511
Qy 143 SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer 162
Db 512 -----GATGGCGCAGCAGCA 526
Qy 163 GluSerAlaGluProAspAspThrLeuGlyGlyGlyTyrAlaTyrValTyrGlyThrAsn 182
Db 527 GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCCAAAAGCTTGTGATTTGGGGAACAGAT 586
Qy 183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr----- 199
Db 587 GTGAATGTGGCAACATGTAAGAGAAATTTTCAGAGATTCCTTCAGTGTTTTACTGATCCT 646
Qy 200 -----ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGln-I1 213
Db 647 CTGGCCAAAGAGAAAGAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACA 706
Qy 213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTy 233
Db 707 CTTGAGAGATTAATATTACAGGAGAGCCATTTTAAATGTGAACCTGCAACACATAAA 766
Qy 233 rAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe 253
Db 767 ATCATTTAGCAAAATCTGTATAGCAGCTCATCTCTACCCACAGGAGTTTATACCAAC 826
Qy 253 uLeuAspThrGluCysGlnGluValAlaIleThrSerLeuLeuPro-----ThrPheGluTy 271
Db 827 CTTTGACATGGCTGTCAATGAGATCTTTTGACCGTTATCTGACTCCATCTTAGAACA 886
Qy 271 sHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPr 291
Db 887 TCAGATTCAAGTCAGACCTTTTAAATCGTTGAAGACAAAGAGTATGAGAACTTGAATCC 946
Qy 291 cSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleI1 311
Db 947 AGAAGACATTTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCAGCTGAT 1006
Qy 311 eProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVa 331
Db 1007 TCCGGAGATCGAGGAGGCTTTTCCAAATGCCAAGTCTGTGCCACACC---ACCCGGGT 1063
Qy 331 lThrValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAl 351
Db 1064 GGAGATAGATCGAGGAGCAATTTGTCAGCCCTGCAGTGT-----GTGCACTGCCACAC 1117
Qy 351 aArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValAr 371
Db 1118 TACCACAGCATGGCAGCTGATCCAACCGATCATCTTCTTCAGAACGAAATGATCAA 1177
Qy 371 gLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCy 391
Db 1178 ACTTCAAGATCTCTTGAAGCATCGCTGCTGGGAGAGACACCTCACATATTGTCCTTTT 1237
Qy 391 sLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVa 411
Db 1238 TGCCCAACATGACCTTGTGTGACAGGTTCAACAGGGGACAGTGAACGTCACAGGCAT 1297
Qy 411 lPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrIy 431
Db 1298 ATATCGAGCAGTACCAATTCGAGTTAATCCAAAGAGTGAGCAACGTTGATCTGTATAA 1357
Qy 431 sThrTyrIleAspCysValHisValLysSerAspArgGlyArgLeuGlnThrGluAs 451
Db 1358 AACCCACATTTGATGTCATTTATTATCGGAAACCGGATGCAAAACGTCCTGCAT----- 1409

Qy 451 pProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSe 471
Db 1410 -----GGCTTGATGAAGACGACAACA 1432
Qy 471 rGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyI1 491
Db 1433 GAAACTTTTTTCAGAGAAACGTGAAATTCCTTAAGGAACCTTTCAGGAAGCCAGATAT 1492
Qy 491 eTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysIy 511
Db 1493 TTATGAGCGGCTTGCTTCAGCCTTGGCTCCAGCATTTATGAACATGAAGATATCAAAA 1552
Qy 511 sGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle-----ProSerGlyAl 529
Db 1553 GGGAACTCTTACTTCAGCTCTTTTGGTGGAAAGGAAGGATTTTCAGTCACACTGGGAGGG 1612
Qy 529 aSerPheArgGlyAspIleAsnValLeuValGlyAspProGlyThrSerLysSerG1 549
Db 1613 TAAATTCGTCTCGATCAACATCTCTCTGTGGGGACCTTGCCACGACGAGTCCCA 1672
Qy 549 nLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySe 569
Db 1673 GCTGCTACAGTATGTGTACAACCTGGTCCCGAGAGCCAGTACAGCTCTGGAAAAGGCTC 1732
Qy 569 rSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValIe 589
Db 1733 CAGTCGGCTCGGCTCACCGCTATGTGATGAAGACCTCGAGACCAGCGCATTTGTCT 1792
Qy 589 uGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLy 609
Db 1793 CCAGACAGTCCCTCGTCTCGATGACATGGGATATGCTGCATCGATAGTTTGACAA 1852
Qy 609 sMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVa 629
Db 1853 AATGAATGAAGACCAAGGCTGTGCTGCATGAGTCTATGGAACAGCAGACACTCTGCTCAT 1912
Qy 629 lAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAs 649
Db 1913 TGCAAGGCTGGGATCATCTGTCACTCAATGCGCGCACCTCTGTCTCTGGCAGCAGCAA 1972
Qy 649 nProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPr 669
Db 1973 TCTTATTGAGTCTCAGTGGAAATCTTAAAAAACCAACCATTTGAAATATATCCAACTACCGCA 2032
Qy 669 oThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAs 689
Db 2033 CACATTGTTGTCAGGTTTGATCTCATTTCTCATGCTAGACCCTCAGGATGAGGCATA 2092
Qy 689 nAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerIy 709
Db 2093 TGACCGGCTGAGTCTATCATCCTCGTTTTCATTGTACTACCAAAAGTGAGGAGCAAGTGA 2152
Qy 709 sGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHiVa 729
Db 2153 GGAGAGTCTCGACATGGCGCTGCTGAAAGACTACATTGCATATGCCATAGTACCAT 2212
Qy 729 lHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetAr 749
Db 2213 CATGCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTTGAGGCTTATGTAACATGAG 2272
Qy 749 gGlnLysGlyAsnPheProGlySerSerLysValIleThrAlaThrProArgGlnIe 769
Db 2273 GAAG-----ATTGGGAGTAGCGGGGATGGTTTCTGCTTACCTCCGACAGCT 2320
Qy 769 uGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValIalG1 789
Db 2321 AGAGTCAATTAATTCCTTAGCAGAGGCCCATGCTTAAAGTAAAGATTTTCAAACAAAGTTGA 2380
Qy 789 uLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaTh 809
Db 2381 AGCAATTGATGTGGAAGAGGCAAAACGCTCCACCGGAGGCTCTGGAAGCAGTCTGCAAC 2440

QY 809 rAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGI 829
Db 2441 TGACCTCGTACTGGCATTTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTC 2500
QY 829 uArgIleArgArgAlaAsnLeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSe 849
Db 2501 TCGTAACCGGAAGAAGAAATTTAGTGAACCATTTGAGAAATTTATT-----TTATC 2551
QY 849 rProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
Db 2552 TAAGGTTAAACACACAGCCCTTAAAGTACCACAGCTGTTTGGAGATATTCGGGGACAGTC 2611
QY 869 rSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGI 889
Db 2612 TGACACAGCAATTTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGA 2671
QY 889 yPheLeuThrValHisGlyAspIleValLys 899
Db 2672 TTTCTTAACAGTACTGGGAAGACTGTCGC 2702

RESULT 15

CR380957_02/c

WPCOMMENT

Sequence split into 13 fragments LOCUS CR380957 Accession CR380957

Fragment Name	Begin	End
CR380957_00	1	110000
CR380957_01	100001	210000
CR380957_02	200001	310000
CR380957_03	300001	410000
CR380957_04	400001	510000
CR380957_05	500001	610000
CR380957_06	600001	710000
CR380957_07	700001	810000
CR380957_08	800001	910000
CR380957_09	900001	1010000
CR380957_10	1000001	1110000
CR380957_11	1100001	1210000
CR380957_12	1200001	1302002

Continuation (3 of 13) of CR380957 from base 200001 (CR380957 Candida glabrata strain CB

Alignment Scores:

Pred. No.:	1,27e-57	Length:	110000
Score:	1819.50	Matches:	397
Percent Similarity:	59.00%	Conservative:	170
Best Local Similarity:	41.31%	Mismatches:	261
Query Match:	39.48%	Indels:	133
DB:	15	Gaps:	21

US-10-768-511-8 (1-901) x CR380957_02 (1-110000)

QY	31	ThrSerProSerPheAspAsnAlaSerProValAlaGlyArgAlaValArgGln	50
Db	21810	ACTTCGGAGATGTCAAATATCCAGTCTCCAGTGATGGGT-----	21769
QY	51	ThrProThrSerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArg	70
Db	21768	---CCTATT-----GATGAGGAGATCTCTCCCGCCCGC	21736
QY	71	SerArgSerArgSerLeuGly-----	Asn 78
Db	21735	AACCGCCCTCAAGCATGGGTCCAAATGGTGATCCATCCAGTCCAGCATTTGTTTTTAAC	21676
QY	79	SerValTySerSerProTyAspAlaGlyThrProGlyThrProGlyThrProValAla	98
Db	21675	TCATCATCATCCCAACACAGGGTGATTTCAACCTGTGCGCCCATCTTCGCCCAATTCAT	21616
QY	99	ThrProValTyAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGly	118
Db	21615	TTCCCA-----TCTTCCTTAACAGACCAATGAGCTCAGATGTTCACTCTCAAGGA	21565
QY	119	ThrProGlnTyLysGlnArg-----	Ser 126
Db	21564	AGAAGGAGCAACAGAAATCACCTTGCAGCCACACAGATCGAGTGTCTTAGGAAGATTTCTCT	21505

QY	127	GluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluPro	146
Db	21504	GACTACGGGAAGCAGTAGACAGGTACTACATAGTCTCTCTCTTAATATGGCGGCTACCT	21445
QY	147	GlyHisArg-----SerProSer	152
Db	21444	TCTCAGAGCACTTAATTTAAGAAATAATATTCAGCCTCAGATCTATCTCTCCACAGA	21385
QY	153	ArgGluProSerAlaAspGlyArg-----	160
Db	21384	AGGATTGTAGACTTCGACTCCAGGTCCAGGTATACAGCACTTCTTCTTCTCTCTTCA	21325
QY	161	---ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyAlaTyValTtp	179
Db	21324	ATGCTCTCGAGCTACCGAACCT-----CTAAGAATATTATTGG	21286
QY	180	GlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu-----	196
Db	21285	GGTACAAATGTGACATACAGAAATGTGCCAATAGCTTCAGGAATTTCTGTATGTATTT	21226
QY	197	---HisAsnTyArg-----SerSerAlaHisAspLeuAsnSerLys-----	209
Db	21225	AAGTACAGTATAGAGAGTTTGTGATGGAGACAGATATTAACCGATGATGAAGCAGAA	21166
QY	210	-----TyrIleGlnIleLeuGluThrValGluArgGluGluAspThrLeu	225
Db	21165	GAAGAACTATATCTATGTGAAGCAACTTAATGAATGAGAGAACTTGGTATCATCAATTTA	21106
QY	226	AsnIleAspMetSerAspIle-----TyrAspHisAspProAspLeuTyAlaLysIle	243
Db	21105	AACTCGATGACGCAATTTGTAGCATTTAAACAAACCGAGGAGTTTATTATCACTA	21046
QY	244	ValArgTyProLeuAspIleProLeuLeuAspThrGluCysGlnGluValAlaThr	263
Db	21045	TTGAACATCCCAAGAGTCATATCAATATGGATCAACCATCAAGACTGTATGGTA	20986
QY	264	SerLeu-----LeuProThrPheGlu---LysHisIle	273
Db	20985	TCTCTGTGTGATAACCAATTAGAACATGATTTAGACGAAATTCAGTCAAGTTTATC	20926
QY	274	GluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAsp	293
Db	20925	AAAAGTAAGACCTTACAAATGTTGAAAATCAAAAAGGTATGAGAGAAATTTGAATCAAATGAT	20866
QY	294	IleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleProGlu	313
Db	20865	ATAGATAAGTTGATAGCCTGAAGGTTGGTTCTTAGAGCGACACCTGTATACCAAG	20806
QY	314	IleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProLeuValThrVal	333
Db	20805	ATGAAAGTAGCTTTTTCAAATGTAATATTTGTGATCACACC---ATGGCTGTTGAAATC	20749
QY	334	ValLysGlyArgValGluGluProThrArgCysGlyLysProGluCysAlaArgAsn	353
Db	20748	GATCGTGTGTAATTCAGGAACCCAGCTAGATGTGAACGTTGTTGTTGTTGTTGTTGTTG	20689
QY	354	AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln	373
Db	20688	TCTATGACCTTAATCATATAGATGCTCTTTCGCTGATAAACAAGTTATTAACTGTCAA	20629
QY	374	GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTy	393
Db	20628	GAACCTCTGATTTGTTCCAGATGACCAACACCTCACTCTCTCTCTCTCTCTCTCTCT	20569
QY	394	AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys	413
Db	20568	GACGAGTTAGTTGATAGTTGCGGTGCTGTCAGAGAAATGAAGTTACAGTACCTCTCAG	20509
QY	414	AlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyLysThrTy	433
Db	20508	TCAATCCCATCAAGCCCAACTCTAGAACACGCTCTTAAATCTTTTATACAAAATAC	20449

```
QY 434 IleAspCysValHisValIysSerAspArgGlyArgLeuGlnThrGluAspProMet 453
Db 20448 ATTGATGTCGTCATGTAAGAGGTCTCGAACAACATAGAAATGGTGT---GATGTTTCA 20392
QY 454 GluMetAspLysGlu-----AsnAspMetTyrAlaGlyTyr 465
Db 20391 ACTATCGAAGAAGAACTATTGCAGAAATAAAGCTAGATAATATGATGTC----- 20344
QY 466 HisGluSerAspThrSerGluAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 485
Db 20343 -----GAAGAAGTAAGACAAATCTCTGATGCTGAAATCGAAGATAAAGCAGGT 20293
QY 486 SerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGlu 505
Db 20292 GCACAAAGCCGATCTTTATGATCTTCTGCGAGATCTATAGCCCAAGTATTATGAA 20233
QY 506 LeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysIle 525
Db 20232 TTAGATGACGCAAAAGGGTATATTACTACAATTATTTGGTGCTACTAACAAACGTTT 20173
QY 526 ProSerGlyAlaSerPheArgGlyAspIleAlaValLeuValGlyAspProGlyThr 545
Db 20172 AAGAAGGGTGGTGGTATAGAGGTGATATTACATTTTGTATGTTGGTGAACCCCTCCACA 20113
QY 546 SerLysSerGlnLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSer 565
Db 20112 TCCAAATCCAGATCTTCAATATGTTTCATAAAATGCTCTAGAGGTGCTTACACATCT 20053
QY 566 GlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArg 585
Db 20052 GGTAAGGGTTCATAGCAGTAGGTCTAACGGATATGTCACAGAGATGTCGATAGTAAA 19993
QY 586 GluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAsp 605
Db 19992 CAATTAGTTTATAGAGAGTGGTCCCTAGTCTTCTGATGGTGGTATCTGTTGTTATGAT 19933
QY 606 GluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 625
Db 19932 GAATTTCGATAAATGAGTGAATCAATAGATCTGTTCTTCAGAAAGTATGGAACAA 19873
QY 626 ThrValSerValAlaLysGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeu 645
Db 19872 AGCATCTCGTGGCTAAAGCTGGTATATATCAACAATGAATGCTAGAGTTCCATTTTA 19813
QY 646 AlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIle 665
Db 19812 GCTAGTGCAAAACCCCTATTGGATCCCGTTATAACCCCTAATCTACCTGTTACTGAAAATAT 19753
QY 666 GlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysPro 685
Db 19752 GATCTTCCACCCCAATTACTGTCAGATTCGACCTTGTATTATATTATTTGGATAAGTT 19693
QY 686 AspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn--- 704
Db 19692 GATGAAGCAGCTGACCGGTGATCTGGCAAAACACTTGACCACTCTTTATCTAGAGCAAG 19633
QY 705 ---TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThr 723
Db 19632 CCTGCACATGTTACTACCGATGATGTTCTCCCAATCGATTTCTTAACACAATACATCAAT 19573
QY 724 TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsn 743
Db 19572 TATGCAACAGAAATGTTTCATCCTTGTAAACAGCAAGCAAAAAAATCAACTGTGTAAG 19513
QY 744 GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
Db 19512 GCTTATGTTGGCATGAGAAAAATGGTGACGATTCAGATCAGTGAAGAAGAAATTACT 19453
QY 764 AlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
Db 19452 GCAACAAACGAGACAGCTTGAAGTATGATTCGTTTATCTGAAGCCCATGCAAGATGAGA 19393
QY 784 PheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
```

```
Db 19392 TTATCAAGCACTGTAGATCTTGAAGACGTCGCTGAGGCTGTCAGATTAATGAATCGGCT 19333
QY 804 LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
Db 19332 ATTAAGATTACGCAACTGATCCAAAGACAGGTAAATCGATATGAATTTGGTTCAAACT 19273
QY 824 GlyValSerAlaSerGluArgIleArgArgAlaAsnLeu-----LeuAlaAla 839
Db 19272 GGTAATCAGTTATACAGCGTAAATTAACAGAGGATTTGGCTAGGGAGATAATACGTATT 19213
QY 840 LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSer 859
Db 19212 CTGAAGAGTATCCAGCTGACTCTATGAT-----TTCAAT 19177
QY 860 GlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLys 879
Db 19176 GAGTTGATCAAGCAATCAATGAACAAGCACACAGACAGAGTAGAACCAAGTCAAGTGTC 19117
QY 880 AsnAlaLeuGlySerLeuGlnGlyGlyPheLeuThrValHisGlyAspIleValLys 899
Db 19116 AGCATTTCAGGTTTACAGCAGGAAGATAAAGTGATTATCTCTGGGTGAAGGTGTTAGA 19057
QY 900 Arg 900
Db 19056 AGA 19054
```

Search completed: December 6, 2005, 12:36:37

Job time : 8562 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 06:12:03 ; Search time 867 Seconds

(without alignments)
6926.053 Million cell updates/sec

Title: US-10-768-511-8

Perfect score: 4609

Sequence: 1 MENNDALDIGAVSSPPSQS.....LGSLLQGGFLTVHGDIVKRV 901

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cn2_1/USPTO spool/US10768511/runat_02122005_102333_18396/app_query.fasta_1.1095
-DB=N Geneseq -QFM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511 @CN_1_1_608 @runat_02122005_102333_18396 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 21.1*

1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*
14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4609	100.0	4348	6 AAI67613	AAI67613 Cell cycl
2	4609	100.0	4348	10 AAD60813	AAD60813 Physcomit
3	2550	55.3	2841	13 ADX53995	Adx53995 Plant ful
4	1883	40.9	3072	8 ABT21208	ABT21208 Aspergill

5	1871.5	40.6	2625	8 ABT19388	ABT19388 Aspergill
6	1840	39.9	1723	12 ADJ39580	ADJ39580 Plant cDN
7	1835.5	39.8	2970	10 ADB69927	ADB69927 C. neofo
8	1833	39.8	2802	13 ADT47760	Adt47760 Bacterial
9	1819	39.5	2881	13 ADS45058	Adsa49058 Bacterial
10	1817.5	39.4	3273	6 ABL65258	ABl65258 lung canc
11	1817.5	39.4	3273	6 ABL65666	ABl65666 lung canc
12	1817.5	39.4	3273	6 ABL65259	ABl65259 lung canc
13	1817.5	39.4	3273	12 ADO19849	ADO19849 Human PRO
14	1817.5	39.4	3273	12 ADO19847	ADO19847 Human PRO
15	1817.5	39.4	3273	13 ADU26059	Adt26059 Breast ca
16	1817.5	39.4	3273	13 ADU05807	Adu05807 Novel bro
17	1816.5	39.4	3395	10 ADJ56481	Adj56481 Frog cDNA
18	1811.5	39.3	3248	13 ACN38801	Acn38801 Tumour-as
19	1811.5	39.3	3248	13 AAF17708	Aaf17708 Human cDN
20	1791	38.9	3394	6 AAS94968	Aas94968 Human DNA
21	1789.5	38.8	2842	3 AAF15678	Aaf15678 Human pro
22	1764	38.3	3362	8 AET20510	ABt20610 Aspergill
23	1764	38.3	4944	8 AET18200	ABt18200 Aspergill
24	1764	38.3	5362	8 AET20014	ABt20014 Aspergill
25	1752.5	38.0	2944	8 ABT18794	ABt18794 Aspergill
26	1752	38.0	2733	6 ABZ32193	ABz32193 Candida a
27	1733.5	37.6	5265	12 ADP83382	ADp83382 Breast ep
28	1690.5	36.7	2692	4 ABL02365	ABl02965 Drosophil
29	1677	36.4	5419	10 ADB63205	ADB69205 C. neofo
30	1675	36.3	3419	10 ADB69566	ADB69566 C. neofo
31	1670.5	36.2	1571	13 ADX36902	Adx36902 Plant ful
32	1634.5	35.5	3178	13 ACN41452	Acn41452 Human dia
33	1584.5	34.4	5065	4 ABL02964	ABl02964 Drosophila
34	1139	24.7	723	6 AAI67610	AAI67610 Cell cycl
35	1139	24.7	723	10 AAD60810	AAf60810 Physcomit
36	1070	23.2	2046	4 AAF57040	AAf57040 DNA seque
37	1021.5	22.2	2001	13 ADT46787	Adt46787 Bacterial
38	1016.5	22.1	2280	13 ADT47813	Adt47813 Bacterial
39	1016	22.0	2821	12 ADQ09277	Adq09277 Human MCM
40	1003	21.8	2445	8 AET20750	ABt20750 Aspergill
41	998	21.7	2427	6 ABL62640	ABl62640 Colon ade
42	998	21.7	2427	6 AEN95188	ABn95188 Gene #168
43	998	21.7	2427	12 ADO19789	ADO19789 Human PRO
44	998	21.7	2427	12 ADO19648	ADO19648 Human PRO
45	998	21.7	2427	13 ADR24507	ADR24507 Breast ca

ALIGNMENTS

RESULT 1
AAI67613
ID AAI67613 standard; cDNA; 4348 BP.
XX
AC AAI67613;
XX
DT 27-FEB-2002 (first entry)
XX
DE Cell cycle protein 2 (CC-2) encoding cDNA.
XX
KW Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;
KW CC-3; environmental stress; ss.
XX
OS Physcomitrella patens.
XX
PN WO200177354-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011294.
XX
PR 07-APR-2000; 2000US-0196001P.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Costa E SilvaO, Bohnert HJ, Van Thiel N, Chen R;
XX Sarria-Millan R;

DR	WPI: 2002-049151/06.	181	ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg	200
DR	P-PSDB; AAG66003.	2094	ACGAATGTTAACTATCCAGATGCTTAGGCGGATTCGTGATTTCTCCAAATATATCGT	2153
PT	Novel Cell Cycle Stress-Related Protein useful for increasing tolerance	201	SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg	220
PT	to environmental stress, is selected from Cell Cycle Proteins 1-3, or	2154	TCGAGTGTCTCATGATCTTAATTCACGTACATCCAGATCATAGAGGAGCTGTGGAGGT	2213
XX	their orthologs.	221	GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr	240
PS	Claim 4; Fig 2B; 90pp; English.	2214	GAGGAGGATACCTTAATATCGACATGTCAGACATTTATGACCATGATCTCTGATCATAC	2273
XX	The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),	241	AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu	260
CC	isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1	2274	GCAAAAATTTGTTGATACCCATCGACATCATCCCTCTGTGGACACTGAGTGTGAGAA	2333
CC	protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSRPs and	261	ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu	280
CC	encoding nucleic acids are useful for increasing tolerance to	2334	GTTCCTACCTCTTTTACTACCAACGTTTGAGNAGCATATTGAGGCCAGACCTTTCAATCTC	2393
CC	environmental stress selected from salinity, drought and temperature, in	281	LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal	300
CC	transgenic plants including monocot and dicot selected from maize, wheat,	2394	AAAGCATCGGTGCATCGCTGAACTCAACCTTCAGATATAGACAAATTCGTTCTGTT	2453
CC	rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,	301	LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys	320
CC	canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,	2454	AAAGGAATGGTTATCCGGTGCAGTTCTATCATCTGAAATTAAGGGGGCTTCTTCAAA	2513
CC	tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,	321	CysLeuValCysGlyHisSerProProLeuValThrValLysGlyArgValGluGlu	340
CC	tea, Salix species, oil palm, coconut, perennial grass and forage crops.	2514	TGTTTAGTGTGTGTCTACCTCGCTCCGTAGTTACAGTTGTTAAAGGGCGGTGAGGAG	2573
CC	The nucleic acid is also useful for identifying organisms e.g. P. patens	341	ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn	360
CC	in a mixed population of microorganisms. The nucleic acids are also	2574	CCAAAGGTGTGAAAAGCCAGAAATGTGCAGACGGAATGCTATGTCCTCTTTATTCAAA	2633
CC	useful for evolutionary and protein structural studies. The proteins and	361	ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro	380
CC	nucleic acids are useful as markers for specific regions of the genome.	2634	CGATGCACTTTTGCAAAATAAGCAGATAGTGCCTCTTCAAGAAACTCCAGATGCCATCTCT	2693
CC	The present sequence represents the cDNA encoding P. patens CC-2 protein	381	GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal	400
XX	SQ Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;	2694	GAAGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACTATGTTGATGCTGTG	2753
Alignment Scores:		401	LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly	420
Pred. No.:	9,31e-263	2754	AAGCTGTGAGATCGTATTGAGGTAAAGGAGTTCCTCAAGGCCATGGCAGTTTCGAGTTGCT	2813
Score:	4609.00	421	ProAsnGlnArgThrLeuArgAlaLeuTyrIleThrTyrIleAspCysValHisValLys	440
Percent Similarity:	100.00%	2814	CCGAATCAACGAAACATTACGAGCATTTGTATAAGACCTTACATCGATTCGCTGCAGTCAAG	2873
Best Local Similarity:	100.00%	441	LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp	460
Query Match:	100.00%	2874	AAGTCTGACGGGTGCATCGCAACTGCAAGATCTTATGGAGATGGATGAAGAGATGAT	2933
DB:	6	461	MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln	480
US-10-768-511-8 (1-901) x AAI67613 (1-4348)		2934	ATGTATGCTGGTATCATGAAAGTGATACCTTCAGAAAGCTGCTAATGAAGCAAGATTCAA	2993
QY 1 MetGluAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20		481	LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla	500
DB 1554 ATGGAAAATAATGATGCATTCGACATTTGGAGCGGTGCTGCCATATCTTCGCAATCT 1613		2994	AAACTTAAGAGCTGTCCAAAGCTCCCGGGCATTTTATGATAGACTTTCAGGTCCGTGCT	3053
QY 21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaIleSer 40		501	ProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly	520
DB 1614 GAAGAGAGTGTCTACGCCATTCGCGAAGTAACATCAGCAGGCTTCGACAAATGCAGCCTCA 1673		3054	CCAAAGCTTTGGAGAGCTTGAAGATTTAAANAAGGGTCTCTTTGCCAGCTCTTTGGTGGG	3113
QY 41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgGly 60		521	LysAlaLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuVal	540
DB 1674 CCCGTGGCGCGCGAGGCGCTACGCGACACCCCTACATCTGCAGTTTCGAGGAGAGGG 1733		3114	AAGGCTAAGAAAATTCATCTCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGT	3173
QY 61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80		541	GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg	560
DB 1734 AGAGAAACGGATTCGCTCGTGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCGTGT 1793				
QY 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100				
DB 1794 TATAGTTCCTCTTACGATCGGGGACTCTCTGGAATCTCTGGAACTCCAGTGGCTACTCCG 1853				
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120				
DB 1854 GTTTACGCTACCCAGTCGGTACACCTATGGTATGCCCATCGTTCCATCGTGGCAGCCA 1913				
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140				
DB 1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCAGGGGAAAGCCCTCTACATCGGAGACGTGCA 1973				
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160				
DB 1974 TCTCAATCCAGAAACCCCGGCGATCGATCTCTTCAAGGGAACCTAGTGTGATGGCGT 2033				
QY 161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGly 180				
DB 2034 CCTCTGAATCTGCTGAGCCAGATGCACACTTTGGGTGGAGATATGCTATGTTTGGGG 2093				

```
Db 3174 GGGGACCCCTGGTACCAAGTAACTCAGCTGCTTCAGTATGTCACAAAGTAGCTCTCGT 3233
QY 561 GlyIleThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db 3234 GGAATCTACACATAGTGGGCGAGAGTTCGGCGGTGGGCTGACAGCGTATGTAAAGAG 3293
QY 581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db 3294 GATCCAGAACTCGAGAGCGGTATTTGGAGAGCGGAGCTTTGGTCTTAGTGTATCGTGG 3353
QY 601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLysGlu 620
Db 3354 ATATGCTGTATCGATGAGTTCGACAAATGCTGTGATAATGCCGAGCATGCTTCATGAG 3413
QY 621 ValMetGluGlnThrValSerValAlaValGlyGlyIleLeuLeuSerLeuAla 640
Db 3414 GTAATGGAGCAAAACGGTATCTGTAGCCAAAGGGGGTATCATTCCTCGCTGAACGCT 3473
QY 641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db 3474 CGGAGCTCTGCTTCGATGTGCAATCTTAGTGGTCCGATACATGCGGCTTCT 3533
QY 661 ValIleAspAsnIleGlnLeuProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db 3534 GTGATTGATAACATCCAGCTTCTCCCACTCTACTTCTTAGATTTGATTTAATTTACTTA 3593
QY 681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db 3594 ATGCTCGACAAACACAGAGCAAAACGATCGTCTCGCCAGGCGATCTCGTGGCTTTA 3653
QY 701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
Db 3654 CACTATGAACATATGAGTTTCAAAGCAGAGCGCTTGAATCTACAAACATTACCGCG 3713
QY 721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
Db 3714 TATATCACCTATGCTGTCAGCATGTACATCTTACATTAAAGTATGATGATGCTGAAGAT 3773
QY 741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
Db 3774 TTGATTAAATGGCTATGTTGAGATGCGCCAAAGGGCAACTTTCTCGAAGCATAAAG 3833
QY 761 ValIleThrAlaThrProGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
Db 3834 GTGTAACGCCACACTCGCCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893
QY 781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db 3894 CGAATCAGATTTTCTGAAGTGGTAGAGAAAGTTGATGTCAGCAGAGCTGTGCGCTTTTA 3953
QY 801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
Db 3954 GAGCTCGCTTTGAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013
QY 821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaLeu 840
Db 4014 ATCAGACTGGAGTGTGCGCAGCGAGCTATTTCGTGCGGCCCACTTGTCTAGTGTCTGT 4073
QY 841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerGlyLeuLysThrSerGln 860
Db 4074 CGAGAGCTTATAGCAGTAAATTTTCACTTGGCAGCTCTCTGCTGGAAGACGATCAG 4133
QY 861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
Db 4134 CTTCTTGAGGATATCCGAGCCAAAGCAGTGTGAGCGTTAGTTTTCAGGAGTATATTAAT 4193
QY 881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
Db 4194 GCTCTGGTAGCTCCCAAGGAGAGGCTTCTTACTGTCTCATGTGATGATGATCAAGAGA 4253
QY 901 Val 901
```

```
Db 4254 GTT 4256
RESULT 2
AAD60813
ID AAD60813 standard; cDNA; 4348 BP.
XX AC AAD60813;
XX 15-JAN-2004 (first entry)
XX Physcomitrella patens cell cycle protein (CC-2) cDNA.
XX Cell cycle stress-related protein; CCSRP; metal; cell cycle protein; CC;
transgenic plant; environmental stress; stress tolerance; salinity;
drought; temperature; chemical; oxidative stress; gene; ss.
XX Physcomitrella patens.
XX FH Key Location/Qualifiers
CDS 147..4259
FT /*tag= a
FT /product= "CC-2 protein"
FT /transl_except= {pos:1404..2816, aa:Gly-Pro}
XX US2003097675-A1.
XX 22-MAY-2003.
XX 06-APR-2001; 2001US-00828062.
XX 07-APR-2000; 2000US-0196001P.
XX (SILV/) COSTA E SILVA O.
PA (BOHN/) BOHNERT H J.
PA (THIE/) THIELEN N V.
PA (CHEN/) CHEN R.
PA (SARR/) SARRIA-MILLAN R.
XX Costa E SilvaO; Bohnert HJ, Thielen NV, Chen R, Sarria-Millan R;
WPI; 2003-765533/72.
DR F-PSDB; AAS39980.
XX Novel cell cycle stress-related protein capable of conferring stress
tolerance such as tolerance towards salinity, drought, temperature,
chemical, pathogens, to plants upon over-expression.
XX Claim 16; Fig 2B; 62pp; English.
XX The invention relates to an isolated cell cycle stress-related protein
(CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
and its orthologues. Recombinant expression vector is useful for
producing a transgenic plant containing a CCSRP coding nucleic acid,
where expression of the nucleic acid in the plant results in increased
tolerance to environmental stress as compared to a wild type variety of
the plant which involves transforming a plant cell with the recombinant
expression vector, generating from the plant cell a transgenic plant with
an increased tolerance to environmental stress as compared to a wild type
variety of the plant. CCSRP is useful for conferring stress tolerance
such as tolerance towards salinity, drought, temperature, metal,
chemical, pathogens and oxidative stress or their combinations to plants.
XX The present sequence is Physcomitrella patens CC-2 cDNA
SQ Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 9,31e-263 Length: 4348
Score: 4609.00 Matches: 901
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
```

US-10-768-511-8 (1-901) x AAD60813 (1-4348)

QY 1 MetGluAAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
Db 1554 ATGGAAATAATGATGCACTTGACATTTGGAGCGGTGCTGCCCATATCCTTCGCAATCT 1613
QY 21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
Db 1614 GAAGAGGTGCTTACGCCATTGCGCAAGTAACATCACCAGAGCTTCGACAAATGCAGCCTCA 1673
QY 41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgGly 60
Db 1674 CCCGTGGCGCGGAGGCCGTACGGCAGACCCCTACATCTGCAGTTTCGAAGGAGAGGG 1733
QY 61 ArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAAsnSerVal 80
Db 1734 AGAGAAACGGATTCCGCTCGTGTAGGAGGAGTCGATCTCGCAGTTTAGGCCAAATTCCTGTT 1793
QY 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db 1794 TATAGTTCCTTACGATCGGGGACTCTCGAACTCCTGGAACTCCAGTGGCTACTCCG 1853
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db 1854 GTTTACGCTACCCAGTCGGTACACTATGGGTACCCCATCGTTCATCGTGGCAGGCCA 1913
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
Db 1914 CAGTACAAACAGCGCAGTGAGCTTGTTCCAGGGGAAAGCCCTCTACATCGGAGACGTGCA 1973
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db 1974 TCTCAATCCAGAAACCCGGGCATCGATCTCCTTCAAGGAACCTAGTGTGATGGCGGT 2033
QY 161 ProSerGluSerAlaGluProAspThrLeuGlyGlyTyrAlaTyrValTyrGly 180
Db 2034 CCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGATATGCTTATGTTTGGGGG 2093
QY 181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
Db 2094 ACGAATGTTAACTTCCAGATGTGCTAGGGCGAATTCGTGCAATTTCTCCACAATATATCGT 2153
QY 201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleGluGluThrValGluArg 220
Db 2154 TCGAGTGTCTCATGATCTTAAATTTCCAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213
QY 221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
Db 2214 GAGGAGGATACTCTAAATATCCACATGTCAGACATTTATGACCATGATCTCTATAC 2273
QY 241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
Db 2274 GCAGAAATTTGTTCCGATACCCTACCATCATCCCTCTGTTGGACACATGAGTGCAGAA 2333
QY 261 ValAlaThrSerLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
Db 2334 GTTGCTACTCTTTACTACCAACGTTTGAGAGACATATTGAGCGCAGACCTTTTCAATCTC 2393
QY 281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db 2394 AAAGCATCGGTGCACATGCTGCATCAACCTTCAGATATAGACAAATTTGTTTCTGTT 2453
QY 301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db 2454 AAAGGAATGTTTATCCGGTGCAGTTCATCATACCTGMAATTAAGGGGGCTTCTTCAAA 2513
QY 321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db 2514 TGTTAGTGTGGTGCACCTCGCTCGCTAGTGTAGTGTGTTAAAGGGCGGGTTCAGGAG 2573
QY 341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db 2574 CCAACAGGTGTGAAAGCCAGATGTGCAGCAGCGGATGCTATGTCTCTTATTCACAT 2633

QY 361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlnThrProAspAlaIlePro 380
Db 2634 CGATGCACTTTTGCATAAATAAGCAGATAGTGCCTCTTCAAGAAACTCCAGATGCCATTCCT 2693
QY 381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db 2694 GAAGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGTTGATGCTGTG 2753
QY 401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db 2754 AAGCTCGAGATCGTATTGAGGTAACAGGAGTTCCTCAAGGCCATCGCAGTTTCGAGTTGCT 2813
QY 421 ProAsnGlnArgThrLeuArgAlaIleuTyrLysThrTyrIleAspCysValHisValLys 440
Db 2814 CCGAATCAACGAACATTAACGAGCATTTGTATAAGACCTACATCATCGATTGCGTCACGCTCAG 2873
QY 441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db 2874 AAGTCTGACAGGGGTGCACCTCAAACTGAAGATCCTATGGAGATGGATAGGAGAAATGAT 2933
QY 461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db 2934 ATGTATGCTGGTATCATGAAGTGATACCTTCAGAGCTGCTAATGAAGCAAGATTCAA 2993
QY 481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db 2994 AAACCTTAAGAGCTGTCCAGCTCCCGGCCATTTATGATAGACTTTCAAGGTTCGTGGCT 3053
QY 501 ProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db 3054 CCAAGACTTTGGGAGCTTGAAGATATTAAGAAGGTCTCTTTGCCACTCTTTGGTGGG 3113
QY 521 LysAlaLysIleProSerGlyValAspPheArgGlyAspIleAsnValLeuLeuVal 540
Db 3114 AAGCTAAGAAAAATTCATCTCGAGCATCTTTCCGAGGTGACATCAATGTTTACTTGT 3173
QY 541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db 3174 GGGGACCTGTGTACCATGTAATCTCAGCTGCTCAGTATGTGCACAAAGATAGCTCCTCGT 3233
QY 561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db 3234 GGAATCTACACTAGTGGCGGAGGAAGTTTCGGCGGTGGGCTGACAGCGTATGTAACGAG 3293
QY 581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db 3294 GATCCAGAACTCCGAGAGCGGTATTGGAGAGCGAGCTTTTGGTTCTTAGTGATCGTGGG 3353
QY 601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db 3354 ATATGCTGTATCGATGAGTTCCGACAAAATGTCTGATATATGCCGGAAGCATGCTTCATGAG 3413
QY 621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db 3414 GTAATGGAGCAACAAACGCTATCTGTAGCCAAAGGGGTATCATTTGCCCTCGCTGAACGCT 3473
QY 641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db 3474 CGGAGCTGTCTCTTGCATGTGCAAAATCCTAGTGGTCCCGATACAATGCGCGCTTCT 3533
QY 661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db 3534 GTGATTTGATAACATCCAGCTTCTCCCACTCTACTTCTTAGATTTTGAATTTAACTACTTA 3593
QY 681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db 3594 ATGCTCGCAACCAACAGCAGCAAAACGATCGTGTCTCGCCAGGCACTCTCGTGGCTTTA 3653
QY 701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
Db 3654 CACTATGAAACTATGAAGTTTCAAGCAGGACGCTTAGATCTTACAACACTTTACCGCG 3713

QY 721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluasp 740
 Db 3714 TATATCACTATGCTCGTCAGCATGTACATCTACATTAAGTGTAGAGCTGCTGAAGAT 3773
 QY 741 LeuIleAenGlyTyrValGluMetArgGlnLysGlyAsnProGlySerSerIlyLys 760
 Db 3774 TTGATTAATGGCTATGTTGAGATGGCCAAAGGCAACTTTCTCGGAAGCAGTAAAG 3833
 QY 761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
 Db 3834 GTGATAACAGCCACACTCGCACTCGAAGATATGATTCGTATCAGTGAAGCCCTAGCT 3893
 QY 781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
 Db 3894 CGAATGAGATTTTCTGAAGTGGTAGAAGAAAGTTGATGTCAGCAGAAAGCTGTGGCCCTTTTA 3953
 QY 801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
 Db 3954 GACGTGCTTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGATCTT 4013
 QY 821 IleThrThrGlyValSerAlaSerGluArgIleArgAlaAsnLeuLeuAlaAlaLeu 840
 Db 4014 ATCAGACTGGAGTGTGGCCAGCAGCGATTCTCGGGCCAACTTGTAGCTGCTCTG 4073
 QY 841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
 Db 4074 CGAGAGCTTATAGCAGATAAAATTTTCACTGGCAGCTCTCTGGCTTGAAGACCACTGAC 4133
 QY 861 LeuLeuGluAspIleArgSerGlnSerValAspValSerLeuGluAspIleLysAsn 880
 Db 4134 CTTCCTGAGGATATCCGAGCCAAAGCAGTGTGGAGCTTGTATTTGAGGATATTAATAAT 4193
 QY 881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
 Db 4194 GCTCTGGGTAGCTCCAGGAGAGAGGCTTCTTACTGTCCATGCTGGTACATGACATGAC 4253
 QY 901 Val 901
 Db 4254 GTT 4256
 RESULT 3
 ADX53995
 ID ADX53995 standard; cDNA; 2841 BP.
 XX ADX53995;
 AC ADX53995;
 DT 21-APR-2005 (first entry)
 DE Plant full length insert polynucleotide seqid 28735.
 KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content; gene; ss.
 XX Unidentified.
 OS
 XX US2004034888-A1.
 PN
 XX 19-FEB-2004.
 PD
 XX 28-APR-2003; 2003US-00425114.
 PF
 XX 06-MAY-1999; 99US-00304517.
 PR
 XX 05-NOV-2001; 2001US-00985678.
 XX (LIUJ/) LIU J.
 PA (ZHOU/) ZHOU Y.
 PA (KOA/) KOVALIC D K.

(SCRE/) SCREEN S E.
 (TAB/) TABASKA J E.
 (CAO/) CAO Y.
 LIU J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 WPI; 2004-180133/17.
 New recombinant DNA construct, useful for improving plant tolerance to cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or pests, for conferring increased resistance to plant disease, or for improving yield.
 Claim 1; SEQ ID NO 28735; 15pp; English.
 The invention describes a recombinant DNA construct comprising a polynucleotide consisting of a sequence encoding an amino acid sequence available in electronic form from the US patent office at ftp.segdata.uspto.gov/sequence.html?DocId:2004034888. The polynucleotide of the invention are also useful in physical arrays of molecules and as plant breeding markers. The recombinant DNA construct is useful for improving plant tolerance to cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or pests, for manipulating growth rate in plant cells by modification of the cell cycle pathway, for conferring increased resistance to plant disease, for producing galactomannan, lignin or plant growth regulators, for increasing the rate of homologous recombination in plants, for improving yield by modification of photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake or by providing improved plant growth and development under at least one stress condition or for modifying seed oil or protein yield and/or content. This sequence represents a plant full length insert polynucleotide that can be used in the recombinant DNA construct of the invention.
 Sequence 2841 BP; 707 A; 722 C; 730 G; 682 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,77e-141 Length: 2841
 Score: 2550.00 Matches: 514
 Percent Similarity: 73.24% Conservative: 143
 Best Local Similarity: 57.30% Mismatches: 176
 Query Match: 55.33% Indels: 64
 DB: 13 Gaps: 17
 US-10-768-511-8 (1-901) x ADX53995 (1-2841)
 QY 20 SerGluGlyValSerThrProLeuProGlnValThrSerProSerPheAsnAlaAla 39
 Db 100 TCTGACGGCGCGCAGCTCCCTCTCCCGCATCTGCGCCC---TATGGTCGGCGTCA 156
 QY 40 SerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArgArgArg 59
 Db 157 AGCCCACTCGCG-----GTTACTACTCTCCCATCTCAGCCACACGCCGCCCC 207
 QY 60 GlyArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAenSer 79
 Db 208 GGC-----GTCGCGCTCGCGCTGCTGCTCC----- 231
 QY 80 ValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThr 99
 Db 232 -----GCTAGTCTTAT---GCTTCGCTCCCTCTCTCGGGGG----- 267
 QY 100 ProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThr 119
 Db 268 -----TTCGACACGCG-----CCGACCTCGCGCGCGCAGC 300
 QY 120 ProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 139
 Db 301 CCG-----TCCGAGAGCTGGGGCT---GGCGCCCTCGGACGCGCGCTCAG 342
 QY 140 ArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGly 159
 Db 343 AACTCGACTGGCGGTTCCG-----CCGACGCGCTCCATCTCCAAATGTCCACCGATGAC 396

QY 160 ---ArgProSerGluSerAlaGluProAspAspThrLeuGlyGly----- 173
Db 397 GTCCCTCATCTCCGAAGCTGGGACACGACCGCGCGCGCGCGTGCAGCC 456
QY 174 GluTyrAlaTyrValTyrGlyThrAsnValAsnIleProAspValLeuArgAlaIleArg 193
Db 457 ACCCGGCTTCGTCTGGGGACCAACATAGCGTGCAGGAGCTCAACGCCGCCATCTC 516
QY 194 ArgPheLeuHisAsnTyrArgSerSerAlaHisAspLeu----- 206
Db 517 CGGTTCCTGGCCACTTCGG---GACCGCGCGACGCTGGCGCGTGCAGCCCGGTCA 573
QY 207 ---AsnSerLysTyrIleGlnIleGluThrValGluArgGlu---GluAspThr 224
Db 574 GACGAGGCAAGTACATCGCGCCATCCACCGCATCTCCGAGCTCGAGGCGGGAGTCG 633
QY 225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244
Db 634 CTCGAGCTCGATCGCACGAGCTGTTGACACGACCCAGACCTCTACGACAGATGTT 693
QY 245 ArgTyrProLeuAspIleProLeuLeuAspThrGluCysGlnGluValAlaThrSer 264
Db 694 CGCTATCCGCTCGAGGTCTGCCATCTTCGACATCGTCTCATGGACCTCGTCGCGCG 753
QY 265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284
Db 754 ATCGAGCGCTCTTCGGAAGCACATCCAGACCGAGGTATCAACCTCAAGTCGTCAT 813
QY 285 HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetVal 304
Db 814 TGCTTGAGGAATCTCAACCATCTGATTTGGAAGATGGTATCTCATCAGGGTATGATA 873
QY 305 IleArgCysSerSerIleProGluLysGlyAlaPheLysCysLeuValCys 324
Db 874 ATTAGATCAGCTCGTATACCGGAGCTCAAGGAGGTGTGTTCCGCTCGCTGTTGT 933
QY 325 GlyHisSerProProLeuValThrValLysGlyArgValGluGluProThrArgCys 344
Db 934 GGTTCCTACTCAGAGCCGCTGTGTTGATGAGAAGAGTAACCTGAACCCACACATTTGT 993
QY 345 GluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPhe 364
Db 994 CAGAAAGAACATGTAAGCCCAAAATTCATGACCTTAGTGCACCAACAGATGCAGATT 1053
QY 365 AlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThr 384
Db 1054 TCAGACAAGCAGATCATAAAGTTGCAGGAAACACCCAGCAGATACCAAGAGTGGCACT 1113
QY 385 ProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAsp 404
Db 1114 CCACATACAGTTAGTGTCTTGATGATGATAGTCTGTTGATGCTGGAAGCCCTGAGAT 1173
QY 405 ArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArg 424
Db 1174 AGGGTTGAGATAACTGGATATACAGAGCTATGAGTATTCGATTTGGACCAACTCAAAG 1233
QY 425 ThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysSerAspArg 444
Db 1234 ACAGTGAAGTCTATATTCAAGACATATATTGATTCCTTCACATAAAGAACAGACACAAG 1293
QY 445 GlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGly 464
Db 1294 TCTAGGCTTCATGTGCGGACACCACTGGATATTGATTAATTCCTAAC----- 1338
QY 465 TyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGlu 484
Db 1339 ---GCTAGCAATCTACTGAAGAGGATTTCTTAGTGATAGGTTGAGAACTATAAAGAG 1395
QY 485 LeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrp 504
Db 1396 CTTTCGAAGTTGCTGATATCTATGAAGATTTGACTAGATCATATTAGTCCAAACATATGG 1455

QY 505 GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyLysAlaLysLys 524
Db 1456 GAGTTGATGATGTCAAAGAGGCTCTCTTTGCCAGCTTTTCGGCGGTAATCCCTTGAG 1515
QY 525 IleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGly 544
Db 1516 CTTCTCTCGAGCTAGTTTCGGGGTGACATCAATATTTTACTTTGTGGGGACCTGGA 1575
QY 545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
Db 1576 ACAAGTAAATCCAGCTTCTCCAGTACATGATAAACTGTCTCCTCGTGGTATCTATAG 1635
QY 565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
Db 1636 AGTGGTAGAAGAGTCTGCTGTTGCTTACTGCTTAATGTTACCAAGACCTGAGACT 1695
QY 585 ArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
Db 1696 GGCGAAACTGTTTAGAAAGTGGACACTTGTGTTGAGTGACAAAGGTGTTGTTGCATA 1755
QY 605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
Db 1756 GATGAGTTTGATAAGATGCTCTGATAATGCCGGAAGCATGTTACACGAGGTGATGAACAG 1815
QY 625 GlnThrValSerValAlaLysGlyLysIleAlaSerLeuAsnAlaArgThrSerVal 644
Db 1816 CAGACAGTATCCATTCGGAAGCTGGATAATGTCATCTTTAAACGCTAGGACATCTGTC 1875
QY 645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
Db 1876 CTGGCATGTGCCATCTCTACTGATCACTGTTACAAATCCAGGCTCTCTGTAATGCACAC 1935
QY 665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
Db 1936 ATCCACTTAGCGCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATCTTGGACAAG 1995
QY 685 ProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
Db 1996 CGGATGAGCAAACTGATAGCGCGCTGGCAAGCATATGTTTCGTTGCAATTTGAGAA 2055
QY 705 TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyr 724
Db 2056 CCAATTTAGAGGAGCTCGAGGCTTGACCTTCGACACTAGTTTCTCTACATAAGCTAT 2115
QY 725 AlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGly 744
Db 2116 GCAAGGAAGTATATTTCAGCCACAGTTATCTGATGAAGCTGCAGAGAGTAACTCGTGC 2175
QY 745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
Db 2176 TATGTCGAGATGAGAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAGTCAACAGCA 2235
QY 765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
Db 2236 ACCGCTAGACAAATAGAGAGTTGATCGTCTCAGCGAAGCATTAGCCGATGCGGTT 2295
QY 785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
Db 2296 TCTGAAGTGTTCGAGGTGCGGAGTGTGTGGAGGCACTTCAGGCTTCTTGAAGTCGCGAT 2355
QY 805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGly 824
Db 2356 CAGCAGTCTCGACGGATCATGCAACTGCTGATGATGATGATGATGATGATGATGATGATG 2415
QY 825 ValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle 844
Db 2416 ATATCCGACGCAAGAGGCGAGGCGGAGAACCTCGTTGCCGCAACCCGTAACCTGATT 2475
QY 845 AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAsp 864
Db 2476 GCGGAGAAATGACAGCTTGGAGGCCCTCG---ATGCGCATGATTGAGTTGCTGGAGAA 2532
QY 865 IleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySer 884


```
Db 2533 CTGAGGAGCAGAGCTCAATGAAATTCATATGACGAACCTCCGGGTGCTCTTGGCACC 2592
Qy 885 LeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
Db 2593 CTGATGACTGAAGCGCGGTGGTTATTCATGAGAGAACAGTGAGGAGATT 2643

RESULT 4
ID ABT21208 standard; DNA; 3072 BP.
XX
AC ABT21208;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene #3566.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS Aspergillus fumigatus.
XX
FN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PP 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
PI WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or contain contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX elicit immune response, as a reagent in assays designed to quantitatively
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This polynucleotide sequence represents one of the essential
XX genes of Aspergillus fumigatus of the invention
```

```
SQ Sequence 3072 BP; 717 A; 847 C; 827 G; 681 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 8e-102 Length: 3072
Score: 1883.00 Matches: 438
Percent Similarity: 57.77% Conservative: 157
Best Local Similarity: 42.52% Mismatches: 278
Query Match: 40.85% Indels: 159
DB: 8 Gaps: 23

US-10-768-511-8 (1-901) x ABT21208 (1-3072)
Qy 12 ValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThr 31
Db 1 ATGTCTTCTCCAGCTTCTTCTCGCGTAGGGGC----- 33
Qy 32 SerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThr 51
Db 34 CGCCAGGCAAGGACACGGCAACCTTCTTCCCGCGCGTCAACCGTCTCTGTCAGTCTC 93
Qy 52 ProThrSerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArgSer 71
Db 94 CAGACCAGTA--GTCCGACTCAAGGGCGGCAGATGAACAGTCACAGCGACGCT-CGT 150
Qy 72 ArgSerArgSerLeuGlyAsnSerVal-----Tyr 81
Db 151 GCTTCAGGCGACTGAGAGGGCGAAGCGGAGTCCCTCGTCATCGCTATGTTCTTCCAG 210
Qy 82 SerSerProTyrAspAlaGlyThrProGly---ThrProGlyThrProValAlaThrPro 100
Db 211 TCCTCCCTTCAAAGGCAGACAGCAGTGGCGAACAACACAGATGTTGCGAATGATGAGCCC 270
Qy 101 ValTyrAlaThrProVal-----GlyThrProMet 110
Db 271 -----AGTTCACCAATCGGAGATCTTCTACTATGATGAAGAGACAGACACCTAGA 324
Qy 111 Gly---ThrProSerPheHisArgGlyThrPro----- 120
Db 325 GGAAATGCTCTTACCATGAGAGACTCTCTCCCGTATATGCTAGCTCCAGTCCA 384
Qy 121 -----GlnTyrIysGlnArgSerGluLeuGlySerGlnGlyIysProLeu 135
Db 385 ACTCGAGTCTCAGAAATCGCCAGTCCCGCGGTTCGACATTCACGAGCAGCAGCGGATTA 444
Qy 136 His-----ArgArgArgSerGlnSer 143
Db 445 TTCGTCGTCGTCAGACCCAGCATCGAGCAACCGTGGCGTCCCGCGGTAGCGACCTT 504
Qy 144 ArgGluProGlyHis---ArgSerProSerArgGlu----- 154
Db 505 CATTCGTGGTGGTGTCTTTCTTAGCCCGAACCGTCCGCGCAGAGATTTTTCGATGCTAAT 564
Qy 155 -----ProSerAlaAspGlyArgProSerGluSerAla-----GluPro 167
Db 565 GGTATGCTCGCAGCAGCGATCCAGTTCGATGCCACCTTCGGAATATCCACCCA 624
Qy 168 Asp-----AspThrLeuGlyGlyGluTyrAla---TyrValTrpGlyThrAsn 182
Db 625 GACACTCTGAGCGCGAGCGCTTGGCGGTAGCTCAACCGGTGTGATTTGGGTATCAAC 684
Qy 183 ValAsnIleProAspValLeuArgAlaIleArgPheLeuHisAsn----- 198
Db 685 ATCTCCATCCAGATTCATGTCGCGATTCAAGAACTTTCTTACAACTTCCAAACAAA 744
Qy 199 TyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrVal 218
Db 745 TACCCTGTGTGGCGCAGAGGGCGCAACCCGAGGATGAGACACGTATATGTTGATCAGCG 804
Qy 219 GluArgGluGlu-----AspThrLeu 225
Db 805 GAGGAACCGGAAATACATCAGCATGTTGACACCATCGCGCAACTGGAGTAACACCTTG 864
Qy 226 AsnIleAspMetSerAspIle-----TyrAspHisAspProAspLeuTyrAlaIle 243
```


QY 375 hrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyraSnt 395
 Db 1016 CACCTGACAGATTCCTGATGCCAGATCTCTCACTCGGTTTCCTGTTGATGATG 1075
 QY 395 hrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaM 415
 Db 1076 AGCTGGTGGATGCTGCAAGGCTGGTATCGGTCGAAGTACCGGTATTTTCGGTGCA 1135
 QY 415 etAlaValArgValGlyProHsnGlnArgThrLeuArgAlaLeuTyryrIleA 435
 Db 1136 ACCCTGTGCGGTTAATCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAG 1195
 QY 435 spCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluM 455
 Db 1196 ATGTTCTTCATGTTTCAGAAATCGATCGCAAGAAGTTGGGTATCGCATCGCAACATCG 1255
 QY 455 etAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaA 475
 Db 1256 AGCAGGAGCTCTCGGAACAGCGGCTGGGATGCGAACAACACAGCTAGGCTCACTGCGG 1315
 QY 475 snGluAlaLysIleGlnLysLysGluLeuSerLysLeuProGlyIleTyrAspArgL 495
 Db 1316 AGGAGGAA-----GAGAAATAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTC 1369
 QY 495 euSerArgSerLeuAlaProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuC 515
 Db 1370 TCTCTCGGTCTTGGCCCCAGCATCTACGAGATGAGCAGCTGAAGAAGGAATCCCTGC 1429
 QY 515 ysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer-----PheArgG 533
 Db 1430 TTCAGTTGTTGGAGGCAACCAACAGACCTTCCAGAGGGTGGTAACCCAGATACCGGTG 1489
 QY 533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
 Db 1490 GAGATATCAATATCTCTCTGTGTGACCATCTACATCCAAAGTCCAGCTTCTTCGTT 1549
 QY 553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValG 573
 Db 1550 ACGTCCATAAAGATTGCCCCCTCGCGGTGTGTATACCAGCGCAAGGGCTCTCTCGGTGTG 1609
 QY 573 lyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA 593
 Db 1610 GTCTTACGGGTACTCTACCCCGATCTCTGAACCCCGCAGATGCTCTCGATCGGGTG 1669
 QY 593 laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA 613
 Db 1670 CTTGGTCTTTTCAGACGCGGTATCTGTTCATCGACGAGTTTCACAAGATGAACGAAT 1729
 QY 613 snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG 633
 Db 1730 CCACTCGGTCCGTTCTGATGAAGTCATGGAAACAACACAGCATATCTATCCCAAGGCAG 1789
 QY 633 lyIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS 653
 Db 1790 GCATTTACTACTTTTGAACGCTAGACACGATCTCTGGCTTCCGCCAATCGATCGGTA 1849
 QY 653 erArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuS 673
 Db 1850 GCAGGTACAATCCCAACTTGGCCGTTCTCTCAAAATATTGACCTTCCGCTACCTTGCTCT 1909
 QY 673 erArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgL 693
 Db 1910 CCCGATTCCGACTTGTATACCTCTGCTGGACCGAGTGGATGAGCAGGAAGATCTCGGC 1969
 QY 693 euAlaArgHisLeuValAlaLeuHisTyrGluAsn-----TyrGluValSerLysGlnA 711
 Db 1970 TCGCTAGCACTTGTCAATATGTAACCTGGGAAGACAGACCTGAGCATCTGCCGAGCAAG 2029
 QY 711 spAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisP 731
 Db 2030 AAATCTTCCGATCGAATTCCTTACAGCTATATCACCTACGCAAGCAAGCAAGTCCATC 2089

QY 731 roThrLeuSerAspGluAlaAlaGluAspLeuIleAenGlyTyrValGluMetArgGlnL 751
 Db 2090 CAGTGCTCACACCGCCCGCGTAAGCCTTGTCCGATGCTTACGTTAAACATGCGTAAGC 2149
 QY 751 ysGlyAsnPheProGlySerLysLysValIleThrAlaThrProArgGlnLeuGluS 771
 Db 2150 TTGGAGATGACATCCGGTCTCTGACCGCGGTATCACCGCTACCACTCGTCAACTGGAGT 2209
 QY 771 erMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysV 791
 Db 2210 CCATGATCCGACTCTCGAAGCGCATCGCGGTATGCGGTATCGCGGAGGTCACTGCGG 2269
 QY 791 alAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspH 811
 Db 2270 ATGATGTGGAGAGCGCTGCGCTGATCCGCTCCGCATCAAGAGCGGCCACTGACT 2329
 QY 811 isAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgI 831
 Db 2330 CTCGACCGGTCTGATCGACATGAGCTTGTTCGAGGGGCACCTAGTCCAGCGAGAGAC 2389
 QY 831 leArgArgAlaAsnLeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSerProG 851
 Db 2390 GCAGCGGGAAGCACTCAAGCGTCCCTTGTCTGTCTGTGTGATCATCTGTGCGCGGTG 2449
 QY 851 lySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerServ 871
 Db 2450 GCGTGTGAGCT-----CGTGGCGGAGGCTTTCAGGATCTTAAGCGAGAACAGCAGCA 2503
 QY 871 alAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheL 891
 Db 2504 TTGAGTGGATGGAGCCAGCTTTCGGATGCGGTTCGAGCGCTGAGGCTGAGGAGCGG 2563
 QY 891 euThrValHisGlyAspIleValLysArg 900
 Db 2564 TGAGTGTGTCGCGAGGTCGCGCGG 2592
 RESULT 6
 ADJ39580
 ID ADJ39580 standard; cdna; 1723 BP.
 XX
 AC ADJ39580;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Plant cdna #580.
 KW Plant; gene; ss; transcription; plant genome augmentation; cereal;
 KW soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
 KW maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
 KW stress tolerance; salt tolerance; cold tolerance; drought tolerance;
 KW plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
 KW antifungal.
 XX
 OS Eukaryota.
 XX
 FN US2004016025-A1.
 XX
 PD 22-JAN-2004.
 XX
 PF 26-SEP-2002; 2002US-00260238.
 XX
 PR 26-SEP-2001; 2001US-0325277P.
 PR 26-SEP-2001; 2001US-0325448P.
 PR 04-APR-2002; 2002US-0370620P.
 XX
 PA (BUDW/) BUDWORTH P.
 PA (MOUG/) MOUGHAMER T.
 PA (BRIG/) BRIGGS S P.
 PA (COOP/) COOPER B.
 PA (GLAZ/) GLAZEBROOK J.
 PA (GOFF/) GOFF S A.
 PA (KATA/) KATAGIRI F.
 PA (KREP/) KREPS J.


```
QY 631 sGlyGlyIleIleAlaSerLeuAenAlaArgThrSerValleuAlaCysAlaAenProSe 651
Db 1066 GGCTGGAATAATTCATCTTTAAATGCCAGAACATCAGTTCCTAGCATGTGCATACTCTTAC 1125
QY 651 rGlySerArgTyrrAenAlaArgLeuSerValIleAspAenIleGlnLeuProProThrLe 671
Db 1126 TGAATCAGTTATATCCAAAGCTCTCTGTGATGACATATCCATCTTCTCCACACT 1185
QY 671 uLeuSer--ArgPheAspLeuIleTyrrLeuMetLeuAspLysProAspGluGlnAenAsp 690
Db 1186 GCTTCTAGAGGTTTGACCTCATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGAT 1245
QY 691 ArgArgLeuAlaArgHisGluValAlaLeuHisTyrrGluAenTyrrGluValSerLysGln 710
Db 1246 AGAGCCCTGGGCTAAGCATATTTCTGTTGCAATTTGAGAAAT----- 1287
QY 711 AspAlaLeuAspLeuGlnThrAlaTyrrIleThrTyrrAlaArgGlnHisValHis 730
Db 1288 -----CCAACTACATTAAGTTATGCAAGGAACATATACAA 1323
QY 731 ProThrLeuSerAspGluAlaGluAspLeuIleAenGlyTyrrValGluMetArgGln 750
Db 1324 CCACAGTTATCTGATGAAGCTGCAGAGAATTAAGTCCGCGCTATGTTGAGATGAGGAAA 1383
QY 751 LysGlyAenPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
Db 1384 AGAGGAACAGCCCTGGTAGCAAGAAAG---ATAACTGCGACAGCTCGACAAATTGAG 1440
QY 771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
Db 1441 AGCTTGATTCGCTCAGTGAGCACTGCCCGAATGCCGATTCCTCGAATATGTTGAAGTA 1500
QY 791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db 1501 CAAGATGTTGTAGAGCGCTTCAGGCTTCTCGAAGTTGCCATGCAGCAATCGCAACTGAT 1560
QY 811 HisAlaThr-GlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluAr 830
Db 1561 CATGCCACTGGGTACAATCGATATGATCTTATCATGACTTGAATATCTGCGAGCGAAA 1620
QY 830 gIleArgAlaAenLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPr 850
Db 1621 GCAGAGGCGG----- 1630
QY 850 oGlySerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSe 870
Db 1631 -----GACCAGTTGCTGGAAGAAATTAGAAGCAGAGCTC 1665
QY 870 rValAspValSerLeuGlnAspIle 878
Db 1666 TATGGAAGTTTCATCTGCATGATGTA 1690
RESULT 7
ADB69927
ID ADB69927 standard; DNA; 2970 BP.
XX
AC
XX
ADB69927;
XX
DT 04-DEC-2003 (first entry)
XX
DE C. neoformans open reading frame SEQ ID NO:2332.
XX
KW ds; gene; fungicide; gene therapy; infection.
XX
OS Cryptococcus neoformans.
XX
PN WO2003052076-A2.
XX
PD 26-JUN-2003.
XX
PF 17-DEC-2002; 2002WO-US040225.
XX
```

```
PR 17-DEC-2001; 2001US-0341261P.
XX
FA (ELIT-) ELITRA PHARM INC.
XX
PI Zamudio C, Eroshkin AM;
XX
DR WPI; 2003-533017/50.
XX
DR P-FSDB; ADB70288.
XX
PT New nucleic acid, useful for preparing a composition for treating an
XX infection caused by Cryptococcus neoformans.
XX
PS Claim 2; SEQ ID NO 2332; 136pp; English.
XX
CC The invention relates to a novel purified or isolated Cryptococcus
CC neoformans nucleic acid molecule comprising a sequence encoding a
CC polypeptide comprising a sequence not given in the specification. A
CC polynucleotide of the invention has fungicide activity, and may have a
CC use in gene therapy. The nucleic acid is useful for preparing a
CC composition for treating an infection caused by Cryptococcus neoformans.
CC The present sequence represents a C. neoformans sequence of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 2970 BP; 711 A; 746 C; 742 G; 771 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.93e-99 Length: 2970
Score: 1835.50 Matches: 409
Percent Similarity: 59.59% Conservative: 147
Best Local Similarity: 43.84% Mismatches: 246
Query Match: 39.82% Indels: 131
DB: Gaps: 20

US-10-768-511-8 (1-901) x ADB69927 (1-2970)
QY 47 AlaValArgGlnThrProThrSerAlaValArg-----ArgGlyArg 61
Db 10 GCTCTGCTGCAGTACAGCAATCGCGATGAAACACCATTTCTTCCCTCCGCTGCTAGT 69
QY 62 GluThrAspSerAlaArgArgArgSerArgSerLeuGlyAsnSerValTyr 81
Db 70 TCCGAGAGATCGACCCCTCGTGCAGCTCGTGCAGGTGACATCCACTCTCTTCCCGCAC 129
QY 82 SerSerPro-----TyrAsp 86
Db 130 TCATCGCGCTCTCTCGCTCGACGACAGCCCAAGTCCCGGCTGACTCTCTTCCGCTCCGT 189
QY 87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValThrProVal 106
Db 190 GCGTCTTCCCTCGT----- 204
QY 107 GlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrrLysGlnArgSer 126
Db 205 -----ATGGGTACCGATAGCTTT-----GGTACTCCGCGTGT- 237
QY 127 GluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluPro 146
Db 238 ---TTCGGATCGGCTGCCCTTACCCTTCCGAGTTGCCCAATCTCAACTT----- 285
QY 147 GlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGlu 166
Db 286 -----GGTCCCGACGCG- 297
QY 167 ProAspAspThrLeuGlyGlyGluTyrrValTTPGlyThrAsnValAsnIlePro 186
Db 298 -----GACGATATCGATGGTATCGTCAAAATTATCTGGGGTACGACAACTCTCTTTGCA 351
QY 187 AspValLeuArgAlaIleArgArgPheLeu----- 196
Db 352 GAGAGCATGAACCTTTTCCCGGATTTTGTAGAGGCTTCAAAACCAAGTACCGTCTGTGT 411
```


Db 2551 TCCAGTGTCCCGTGGATCATGCACAGTTCGACAGAAATTGTAGGAGCTGGAAGAGAG 2610
 Qy 889 GlyPheLeuThrValHisGlyAspIleValLysArgVal 901
 Db 2611 AGCATAGTCAAGGTCAAGGTGAAGAAAGAGAGAGATT 2649
 RESULT 8
 ADT47760
 ID ADT47760 standard; cDNA; 2802 BP.
 XX AC
 XX ADT47760;
 XX
 XX 02-DEC-2004 (first entry)
 XX
 XX Bacterial polynucleotide #22511.
 DE Recombinant DNA construct; transformed plant; improved plant property;
 XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX
 XX Bacteria.
 XX
 XX US2003233675-A1.
 PN
 XX 18-DEC-2003.
 XX
 XX 20-FEB-2003; 2003US-00369493.
 XX
 XX 21-FEB-2002; 2002US-0360039P.
 PR
 XX (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 PI WPI; 2004-061375/06.
 XX
 XX New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 XX Claim 1; SEQ ID NO 46198; 122pp; English.
 PS
 XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition. This sequence represents a bacterial polynucleotide used in
 CC production. Improved lignin production or improved galactomannan
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic

CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX Sequence 2802 BP; 887 A; 541 C; 579 G; 795 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 6.48e-99 Length: 2802
 Score: 1833.00 Matches: 393
 Percent Similarity: 59.36% Conservative: 178
 Best Local Similarity: 40.85% Mismatches: 281
 Query Match: 39.77% Indels: 110
 DB: 13 Gaps: 14
 US-10-768-511-8 (1-901) x ADT47760 (1-2802)
 Qy 2 GluAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
 Db 31 GAGGATAATAACTCC-----AGCTCCCGCTGTGTCCCTAAATCCTGAT 72
 Qy 22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaSerPro 41
 Db 73 TCTGTCCA-----CCACAGCTTTCTTCCCGAGCTCTATTTATAGCTCTCTTCA 123
 Qy 42 ValAlaGly----- 44
 Db 124 TCACAAGGTGATATCTATGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGGGAAC 183
 Qy 45 ---ArgArgAlaValArgGlnThrPro-----ThrSerAlaValArgArgGly 60
 Db 184 ATCAGAGCTGCTATAGGTCTCTCCACTAAATTTTCCATCTTCTTCCCAAGACAAAAT 243
 Qy 61 ArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
 Db 244 TCCGATGTTTCCAAATCTCAAGGCAGACAGGCGAATTCGTTCTTTCGCCGCGCTCT 303
 Qy 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
 Db 304 GGAAGGCTAGATATCATCTCTGATCTGAGAAGTGTAGAGCAGCTCTCTTCTTCC--- 360
 Qy 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
 Db 360 ----- 360
 Qy 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
 Db 361 -----TCTTCTTTAGCGGTAAATGGTCAAAACCGGTGTACATCGGAAGA 405
 Qy 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
 Db 406 AATGATATTCATACATCTGATTTATCTCTCCAGAGAAATGTGGATTTTGATACTAGA 465
 Qy 161 -----ProSerGluSerAlaGlu 166
 Db 466 TCCGGCGTGAATACTTTGGATACTTCTTCTTCTGCTCTCTCCATCGAAGCCAGTGA 525
 Qy 167 ProAspAspThrLeuGlyGlyValTyrAlaTyrValTyrGlyThrAsnValAsnIlePro 186
 Db 526 CCC-----TTGAGAATAATTTGGGTACCAACGTCAGTATCCAG 564
 Qy 187 AspValLeuArgAlaIleArgPheLeuHisAsnTyrArg----- 200
 Db 565 GAATGTACAACTAATTTTCGTAATTTTGTGATGCTTTCAGATATAAATTCGTAATA 624
 Qy 201 -----SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
 Db 625 TTGGATGAAGGGAGGAATTCATTAAACAATACTACCGACGAAGAATACTATATCAAG 684
 Qy 213 IleIleGluGluThrValGluArgGluAspThrLeuAsnIleAspMetSerAspIle 232
 Db 685 CAGCTTAATGAATGAGAGAACTTGGTACCTCTTAATTTAAACTTGGGTAGAACTTA 744
 Qy 233 -----TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
 Db 745 CTTCCTTACAGCAACACAGAGACTTATACCATCACTATTAAATTTATCTCTCAAGAGTG 804

QY	251	IleProLeuLeuAspThrGluCyysGlnGluValAlaThrSerLeu	----- 265
DB	805	ATTCTATCATGGATCAAACTCATCAAGGACTGTATGGTTTCC	TTAATAGTAGACAACAAT 864
QY	266	-----LeuProThrPheGlu-----LysHisIleGluAlaArgProPheAenLeu	280
DB	865	TTGGATTACGATTTTAGATGAATAGACACCAAGTTTATAAAGTAAGCGCTTACAAATGTA	924
QY	281	LysAlaSerValHisMetArgGluLeuAenProSerAspIleAspLysLeuValSerVal	300
DB	925	GGTCTCTGTAAGGTATGCGGCAATTGAATCCAAATGATATTGATAAATGATAAACTTA	984
QY	301	LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys	320
DB	985	AAGGCTCTGTCTCGATCAATCCCGTTTCCCTGATGAAGGTAGCGTTTTCAAA	1044
QY	321	CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu	340
DB	1045	TGCACGCTCGCATATACA-----AGGCGAGTGGAAATTGATGAGGAGTTATCAAGAG	1101
QY	341	ProThrArgCysGluLysProGluCyAlaAlaArgAenAlaMetSerLeuIleHisAen	360
DB	1102	CCCCTAGGTGTGAACGTATTGATTGAATGAACCCCAATTCATGTCTCACTGATTCACAAC	1161
QY	361	ArgCysThrPheAlaAenLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro	380
DB	1162	AGGTGTTCAATTCGCAGATAAACCAAGTCATTAAGTTACAGGAAATCCAGAACTTGTGCCT	1221
QY	381	GluClyGluThrProHisThrValSerMetCysLeuTyrAenThrMetValAspAlaVal	400
DB	1222	GATGGACAAACGCTCACTCTATCTCATATGTTTTACGATGAATTAGTGGATCTTGC	1281
QY	401	LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetalValArgValGly	420
DB	1282	AGGCGGGCGATCGTATTGAGTGTACTGGCAGCTTTCAGGTCATCCCATTTAGAGCTAAT	1341
QY	421	ProAenGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys	440
DB	1342	TCCAGGCACACGCGTACTAAAGTCGTTGTATAAACATACGTCGATGGTCCACGTTAAA	1401
QY	441	LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp	460
DB	1402	AAAGTTTCAGATAAGAGTTAGAGCTCGATATCTTACTATTGAACCAAGAAATTAATCCAG	1461
QY	461	MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAenGluAlaLysIleGln	480
DB	1462	AAACAAGGTAGATCATACAGAGGTCGAAGAGTAAGACAAATTAATCTGATCAGGATTTAGCA	1521
QY	481	LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgIleuSerArgSerLeuAla	500
DB	1522	AAAATTCGCAAGTTGCGGCAAGAGAAGATTATACAGTTTATTAGCCCGCTCTATTGGCC	1581
QY	501	ProSerIleTyrGluLeuGluAspIleLysGlyLeuLeuCysGlnLeuPheGlyGly	520
DB	1582	CCAGTATTACAGCTAGAGAAGTGTCAAGAAGGGTATATTACTTCAGCTATTTGGCGGC	1641
QY	521	LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAenValLeuLeuVal	540
DB	1642	ACGAATAAAACCTTTACCAGGGTGGCGGTTATAGAGGTGCATATAATATTATTACTTTGT	1701
QY	541	GlyAspProGlyThrSerLysSerGlnLeuGlnTyrValHisLysIleAlaProArg	560
DB	1702	GGGGATCTCTTACTTCCAAATCGCAAAATTTGCAATACGTTCAAAATTAATCTCTCTGT	1761
QY	561	GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys	580
DB	1762	GGTGTGTATATCTTCGGGTAAAGGTTCATCTCGCGTTGGTTAACTGCTTATATTACAAG	1821
QY	581	AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly	600
DB	1822	GATGTCGACACAAACACTTGTATTGAAAGTGGTGATTAAGTATTGCTGTATGCTGATGAGGT	1881

Qy	601	IleCysCysIleAepGluPheAepLysMetSerAspAsnAlaArgSerMetLeuHisGlu	620
Db	1882	GTCTTGTGTTATGACGAGTGTGATAAATGAGTGATTTCTACAAGGTCGGTCTTGCACGAA	1941
Qy	621	ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleLeuLeuSerLeuAsnAla	640
Db	1942	GTCATGGAACGACGACTATTTCAATCGCAAAAGCGGGAATTATCAACAACACTTAATGCC	2001
Qy	641	ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer	660
Db	2002	AGAAAGTCTATTTTGGCCAGTGCTAACCCCAATTGGTTTCAGCTACAACCCCAATTTGCCCT	2061
Qy	661	ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu	680
Db	2062	GTGACTGAAATATTTGATCTACCCGCCCACTACTCTTCGAGATTTCGATTCGGTCTATCTTT	2121
Qy	681	MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu	700
Db	2122	GTCTCTGTATAGGTTGATGAGAAATAATGACGAGAACTAGCCAAACACTTAACAAATCTT	2181
Qy	701	HisTyrGluAsn-----TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu	718
Db	2182	TACCTGGAAGATACGCCGACATATTTCTCAACAGCAGCTGCTACAGTTTGAATTTTATA	2241
Qy	719	ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla	738
Db	2242	ACGATGTATATTAGTTATGCAAGAGGACATACACCCCAATAATACCCGAGCGCGCTAAG	2301
Qy	739	GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer	758
Db	2302	ACTGAGCTGTGTCTGCTTATGTAGGAATGAGAAGATGGGTGAGCATTCGAGATCCGAT	2361
Qy	759	LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla	778
Db	2362	GAGAAGAGAAATCACAGCTACCCACAAGACAACCTTGAAAGTATGATTCGTTGGCTGAGCG	2421
Qy	779	LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg	798
Db	2422	CACGCCAAATGAAATTGAAAACGCTCGTAGAGCTGGAGGATGTTCAAGAAAGCCGTTAGA	2481
Qy	799	LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet	818
Db	2482	TTAATTAGATCAGCCCATAAAGATTATGCAACAGACCCCTAAACCGGTAAATCGACATG	2541
Qy	819	AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuAla	838
Db	2542	AATTTAGTTTCAACAGGTAATCAGTTATTCAGAGAAACCTACAGGAGGATTTGTCAAGG	2601
Qy	839	AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThr	858
Db	2602	GAAATTATGAATGTTTGAAGGATCAG-----GCATCAGACTCAATGTCTATTC	2649
Qy	859	SerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIle	878
Db	2650	AATGAGCTTATAAACAATAATCAATGAACACTCTCAAGATGAGTGTAGTCTTCTGATATC	2708
Qy	879	LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal	898
Db	2710	CAGGAAGCCCTGTCAAGATTGCAACAGAGGACAGGTCATTTGCTTGGCGAGGGTGA	2769
Qy	899	LysArg 900	
Db	2770	AGGAGA 2775	
RESULT 9			
ID	ADS49058		
XX	ADS49058 standard; cDNA; 2881 BP.		
AC	ADS49058;		
XX	02-DEC-2004 (first entry)		
XX	Bacterial polynucleotide #3801.		

XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX Bacteria.
 OS US2003233675-A1.
 XX
 PN 18-DEC-2003.
 XX
 PD 20-FEB-2003; 2003US-00369493.
 XX
 PP 21-FEB-2002; 2002US-0360039P.
 XX
 PR (CAO/Y) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 DR
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 PS Claim 1; SEQ ID NO 27488; 122pp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or by
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 2881 BP; 678 A; 858 C; 746 G; 599 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4.5e-98 Length: 2881
 Score: 1819.00 Matches: 419
 Percent Similarity: 58.17% Conservative: 161
 Best Local Similarity: 42.03% Mismatches: 249
 Query Match: 39.47% Indels: 169
 DB: 13 Gaps: 24
 US-10-768-511-8 (1-901) x ADS49058 (1-2881)
 QY 24 SerThrProLeuProGlnValThrSerProSerPheAsp----- 36

Db 1 AGCTCGCTTTGCCGACGACAGACACCAGAGCGATGCTGACCGACTCCTCGCGCA 60
 QY 37 AsnAlaSerProValAlaGlyArgArgAlaValArgGlnThr----- 51
 Db 61 AACGGACGGAGCCAGTTAAATTGGGGGTATGGCTGTCAACCAAAAAAAGGAC 120
 QY 52 -----ProThrSerAlaValArgArgGlyArgGluThrAspSerAla 66
 Db 121 TACACTGAATGCTGCACCACTCCGCA-----GACTCCTCT 156
 QY 67 ArgArgArgSerArgSerArgSerLeuGlyAen-SerValTyrSerSerProTyr-- 85
 Db 157 CCCATCGCTATGCTCCAGTTCTAGCCCGGCGCTCAGCTTATCAACAGTCGATCTT 216
 QY 86 -----AspAlaGlyThrProGlyThrProGlyThrProValAlaThrProVa 101
 Db 217 CGTAGCGAAAGCAGTCAGCTCTTCGTCAGCTCGCAAGGTCGTTGCCGTCGATCCG 276
 QY 101 lTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGl 121
 Db 277 CGAGCGACATCAACGGGAC-CCTCTTCGTACTCCCGCC-----CA 317
 QY 121 nTyrLysGlnArgSerGluLeuGlySerGlnGlyProLeuHisArgArgArgSe 141
 Db 318 GATTCCCGCGCGAATCATCTCGAGAGCGCGGTGTGTGATCCGC----- 363
 QY 141 rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr 161
 Db 364 -----GATGCCCTGGCTCCGATGCCAATC 389
 QY 161 oSerGluSerAlaGluProAsp-----AspThrLeuGlyGlyGlu---TyrAl 176
 Db 390 GTTTGTCCCAATAACCCCAATACCTCTGAGGCGAGATGCCCTTGGTGACAAAGC 449
 QY 176 aTyrValTyrGlyThrAsnValAsnLeuProAspValLeuArgAlaLeuArgPheLe 196
 Db 450 TCTTTGTTGGGTACAAATTTCTCTTGAGATTCGTTCTCCGGGTTCAGGACTTTCT 509
 QY 196 uHisAsn-----TyrArgSerSerAlaHis----- 204
 Db 510 CAGGAACCTTACCAGGAAATACCGGATGTGGCGAGATGGCGGACGAGCGAGACTAT 569
 QY 205 -----AspLeuSerSerLys---TyrIleGlnIleleGluGluThrValGluAr 220
 Db 570 TGGCCATCCGGATCGGATTCGAAGCCATATTTGGGAGGACTGGAGATATGCTGCTCT 629
 QY 220 gGluGluAspThrLeuAsnLeuAspMetSerAspIle-----TyrAspHisAspProAs 238
 Db 530 TGGTACCACAGCTGTATCTGGATCTCCGCGACTCAAGTCTACCCTCCGCGCACCTTGA 689
 QY 238 pLeuTyrAlaLysIleValArgTyrProLeuAspIleleProLeuLeuAspThr----- 256
 Db 690 GTTATGGCATCAGGCCCAACACTATCTTACCGAAATCATCCCGGTTCATGATCAATGCGT 749
 QY 257 ----GluCys-----GlnGluValAlaThr----- 263
 Db 750 ACATGACTGTATGATGGAGTTGGCGCAAAAGGAATGGCCAGCAAGAGCATCCCAAA 809
 QY 263 ----- 263
 Db 810 CTCTAGGACCGCTCCAGTGCCTCACAAAGCTCGGAGCGCAACTTCCCAAGCTCTGAAG 869
 QY 264 -----SerLeuLeuProThrPheGluLysHi 272
 Db 870 GAGCGAGGAACCGCCCGCGCTGCGCAAAACAGCGCGCCCAATTTGAGGACCA 929
 QY 272 sIleGlu-----AlaArgProPheAsnLeuLysAlaSerValHisMetAr 287
 Db 930 AGTCAGCCAGATGGCGTATGTGGTTCGCCCGTTCGAGAGATCATCAACACTCCG 989
 QY 287 gGluLeuAsnProSerAspIleAspLysValSerValLysGlyMetValIleArgCy 307

Db 990 TGACCTCAACCCATCTGCATGACACAGCTGTGTCTCATCAAGGCTTGTCTCATCAGAAC 1049
Qy 307 sSerIlelleProGluIleValGlyAlaPhePheLysCysLeuValCysGlyHis 327
Db 1050 CACGGCGTATCCCGACATGAGATGATATTTTCAAGTCAGTGTGTGTGTCATTC 1109
Qy 327 rProLeuValThrValVal-----LysGlyArgValGluGluProThrArgCysG1 345
Db 1110 T-----ATAACGGTTTCAGTTAGTGTGCAAGATTCTGTGAGCCTACAGAGTGTCC 1160
Qy 345 uLysProGluCysAlaAArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAl 365
Db 1161 TCGGGCTCGCTCGCGCTCCCAAGAACTCAATGAGATCAATCCCAACAGATGTGCTTCGA 1220
Qy 365 aAsnLysGlnIleValArgLeuGlnThrProAspAlaIleProGluGlyGluThrPr 385
Db 1221 AGACAGAGCAGTCACTGCAAGAACTCGGATACGTGCTGTGTCAGACACC 1280
Qy 385 oHisThrValSerMetCysLeuTyraAsnThrMetValAspAlaValLysProGlyAspAr 405
Db 1281 CCACTCGGTATCCGTTGGTCTATAACGAGCTTGTGGACTTTTGCAGAGCGCGTGACAG 1340
Qy 405 gIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgTh 425
Db 1341 AGTTGAGCTCACTGTGTATCTTTAAAGTCAACCCCGCTTCGTGTCAACCCGCGCATGAGGAC 1400
Qy 425 rLeuArgAlaLeuTyLysThrTyLysThrTyLysThrTyLysValHisValLysSerAspArg1 445
Db 1401 GGTCAAGACGGTGCACAGACATATGTTGATGTTGCTTCAATGTACAAAGGTGCACCGCAA 1460
Qy 445 yArgLeuGlnThrGluAspProMetGluMetAsp-----LysGluAsnAspMetTyra1 463
Db 1461 GAGGATGGTCTCT---GATCCCTCGACCTCGACCTTTCGAGGAGGAGGAGGAGCCATGC 1517
Qy 463 aGlyTyHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLy 483
Db 1518 TAATGCCACGAGCATGAGCAAGTCCGAAAGGTCTCTCTGATGAGGAAGAGGATCAA 1577
Qy 483 sGluLeuSerLysLeuProGlyLysLeuSerArgLeuSerArgSerLeuAlaProSer11 503
Db 1578 GGAACCTGCGCTCGACCGATATCTACGATCTCTTTTTCAGCTTCTTGGCTCCATCAAT 1637
Qy 503 eTPGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLy 523
Db 1638 TTACGAGATGATGATGTGAGAGGGTATCTTCTCACTGTTCCGAGGACAGCAAA 1697
Qy 523 sLysIleProSerGlyAlaSer-----PheArgGlyAspIleAsnValLeuLeuValG1 541
Db 1698 GACCTTTGAGAGGTGGAAGCCCCCAATATACCGTGGAGACATCAACGTTCTCTCTGTGG 1757
Qy 541 yAspProGlyThrSerLysSerGlnLeuLeuGlnTyValHisLysIleAlaProArgG1 561
Db 1758 TGACCCATCCATTCCAAGTCCAGCTTCTCTCTATGTTTCAGAAATGCCCCAGAGG 1817
Qy 561 yIleTyThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyValThrLysAs 581
Db 1818 TGTGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1877
Qy 581 pProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly11 601
Db 1878 TCCCGAGTCCCGCAACTCGTTCTAGAGTCCGGTGTCTCTCTTATCCGACGCGGTGT 1937
Qy 601 eCysGlyAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluVa 621
Db 1938 CTGTGTATCGACGAGTTCGACAGATGAAAGCAATCGACCCGCTCGGTCTTCCAGAGT 1997
Qy 621 lMetGluGlnThrValSerValAlaLysGlyGlyIleAlaSerLeuAsnAlaAr 641
Db 1998 CATGGAACAAACACCGGTGTCGTCGCAAGCGCGGATATATCACCCTCTCAATGCCCG 2057
Qy 641 gThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyAsnAlaArgLeuSerVa 661
Db 2058 AACATCAATCTGGGTCTGCCAACCCATTTGTTAGTAGATACAAACCTCTATCAT 2117

Qy 661 lIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyLeuMe 681
Db 2118 GCGCAAAATATTGACCTCCGCCAACCTTCTATCCGTTTCGATCTGGTCTACCTCAT 2177
Qy 681 tLeuAspLysProAspGluGlnAsnAspArgLeuAlaArgHisLeuValAlaLeuHi 701
Db 2178 CCTCATCTGTGTGATGAGAAGATGATCAGGCTGTGGCCGCCCATTTGCTCTCCATGTA 2237
Qy 701 sTyGluAsnTyGluValSerLysGlnAspAlaLeuAspLeu----- 715
Db 2238 TCTCGAAGACAGCCCGAGAGTGGCCAGCAGCGCCAAACGACGTCCTTGTGAGTTTTTCCCC 2297
Qy 715 ----- 715
Db 2298 GTCTGTCAATCCACAGAGATGAACCCCAAACTAACCAACTCCCTCAAAAAACAGC 2357
Qy 716 -----GlnThr-LeuThrAlaTyIleThrTyAlaArgGlnHisValHisProThrLeuS 734
Db 2358 CCGTCGAGTTTCTAACCTCTCTACATCTCTACGCGCGCTCGCACATCCACCCGCGCTCA 2417
Qy 734 erAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnP 754
Db 2418 CCGCCGAGCCCGTCGCGAGCTCGTCGATCGCTAGTCGATGCGCAAGCTGGGCAAG 2477
Qy 754 heProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleA 774
Db 2478 ACGTGGCGCGCCCGAGAGCGCATTCAGGCCACCAACACGTCAGCTCGAGTCCATGATTC 2537
Qy 774 rGleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaA 794
Db 2538 GCTTAGCCGAGCGCAGCAGCAAGATGCGTTTATCTCAACCCGTCAGCGCGAGCGCTCC 2597
Qy 794 lAGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrG 814
Db 2598 GCGAGCCGCTGCGCTCTCAAGTCGCGCTCAAGACCCCGCCACCGAC---AGCCAGS 2654
Qy 814 lYThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgA 834
Db 2655 GCGCATCGACATGACGTGTTTGACAGAGGCGCAGAGCGCGCGGAGCGCGGCGCAAGG 2714
Qy 834 lAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerS 854
Db 2715 CCGACATGAAGGATGCGGTATTCGCGCTCTGATGAGATGACACCGCGCGGACAGGTA- 2773
Qy 854 erGlyLeuLysThrSerGlnLeuGluAspIleArgSerGlnSerSerValAspValS 874
Db 2774 -----GTTAGGTACTCGAGGTGGCGAGACGGTGTGGCGAGGGGGGGGTGTCAGGTTG 2828
Qy 874 erLeuGlnLysIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
Db 2829 AGCCGCTGAGTTTGGGAGGTTATGAGGCTCTCGAGATGAGGGC 2875

RESULT 10

ABL65258
ID ABL65258 standard; DNA; 3273 BP.

XX ABL65258;

AC ABL65258;

XX 15-MAY-2002 (first entry)

DT 15-MAY-2002 (first entry)

XX Lung cancer related gene sequence SEQ ID NO:3595.

DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.

XX Homo sapiens.

OS Homo sapiens.

XX WO200194629-A2.

PN 13-DEC-2001.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US010838.
 XX 05-JUN-2000; 2000US-0209473P.
 XX 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234059P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 28-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237318P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 01-NOV-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX (AVAL-) AVALON PHARM.
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 3595; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell

CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour

XX SQ Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

XX Alignment Scores:

Pred. No.: 6.4e-98 Length: 3273
 Score: 1817.50 Matches: 406
 Percent Similarity: 57.78% Conservative: 151
 Best Local Similarity: 42.12% Mismatches: 256
 Query Match: 39.43% Indels: 151
 DB: 6 Gaps: 18

US-10-768-511-8 (1-901) x ABL65258 (1-3273)

QY 38 AlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArg 57
 DB 21 GCGACGTGTCGCGAAGCGCGCGC-----CTTCCACGTAAACGCG---CGC 62
 QY 58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgArg-----72
 DB 63 CGCGCGGCGGCGGCGTGGCGCGGAGCGCGGAAACGTCGCGCTGCGGAGCGAGCAG 122
 QY 73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
 DB 123 GGAAGCCGCGAGGCGGCGC---CCGCGCGGAGCTGTCTTGTGCGCGAGGTACTCCGAGC 179
 QY 92 ThrProGlyThrProValAlaThrProVal-----TyrAlaThr 104
 DB 180 ACTATGTCTCCCGCGGTGCGACCCGAGCGCGCGCGCGCGCGCGTGAAGGCGCACC 239
 QY 105 ProValGlyThrPro-----109
 DB 240 CCCGCCAGACGCGCTCGGAGTGAGGATGCCAGGTCACTCCCTCTCAGAGACGTAGAGC 299
 QY 109 -----109
 DB 300 GAGGATTCACCTCCACGCGGAGTTGCGCGGATGCCAACCTCCCTGGAGTGGACCTG 359
 QY 109 -----109
 DB 360 CAGACACTGCTGCGCAGGACGTGCTTTTCCAGCCCTCCCAATGCAATTCCTCAGCT 419
 QY 110 -----MetGlyThrProSerPheHis---116
 DB 420 ATCCCTCTTGACTTTGATTTGATTTACCACTACATACGCGCACTCCCGACTCTCGGTA 479
 QY 117 -----ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
 DB 480 GAGGGAACCCCAAGAGTGGTGTAGGGGCACT---GTGAGACAGAGGCGCTGACCTG 536
 QY 129 GlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluProGlyHis 148
 DB 537 GGCTCTGCAGAGAGGCGCTCAAGTGCATCTGCACTCT-----575
 QY 149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
 DB 576 -----GACGGGCGAGCAGCAGAGATATAGTGGCAAGT 608
 QY 169 AspThrLeuGlyGlyGluTyrAlaTyrValTgGlyThrAsnValAsnIleProAspVal 188
 DB 609 GAGCAGTCTTAGGCCCAAAACTTGTGATCTGGGGAACAGATGTAATGTGGCAGATGC 668
 QY 189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----Arg 200
 DB 669 AAAGAAACTTTCAGAGATTTCTTCAGCGTTTATTGACCTCTCGCTAAAGAGAGAA 728
 QY 201 SerSerAlaHisAppLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
 DB 729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGAGAGTAATGTT 788
 QY 220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239

```
Db 789 ATTGGTGGCAATTTTAAATGTGAACATCAAAATCAATTTGACAAAAATTGG 848
Qy 240 TyrAlaValIleValArgTyrProLeuAspIleIleProLeuAspThrGluCysGln 259
Db 849 TAAGACACTCATCTCTTACCAACAGGAAGTATTCCAACTTTTGACATGGCTGCAAT 908
Qy 260 GluValAlaThrSerLeuLeuPro-----ThrPheGluLysHisIleGluAlaArgPro 277
Db 909 GAAATCTCTTTGACCGTTACCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968
Qy 278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db 969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAGACATTCACCAAGCTC 1028
Qy 298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db 1029 ATCACCATTGAGGGCATGGTATGATCAGACATCCAGCTGATTTCCAGATGAGAGGGCC 1088
Qy 318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValLysGlyArg 337
Db 1089 TTCTTCCAGTGCACAGTGTGCCCCACAGC---ACCCGGGTGGAGATGGACCGCGCGC 1145
Qy 338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db 1146 ATTGCAGAGCCAGTGTGCGGGCGC-----TGCCACACACACCCACAGCATGGCACTC 1199
Qy 358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db 1200 ATCCACACACCGTCTCTCTGTGACAGCAGATGATCAAGCTTCAGGAGTCTCCGAA 1259
Qy 378 AlaIleProGluGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db 1260 GACATGCTCGAGGGGACACACACACAGTTATCTCTGTGCTCACATGATCTCGTT 1319
Qy 398 AspAlaValIleProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db 1320 GACAAGGTCACGCTCGGGACAGAGTGAATGTTACAGGCATCTATCAGAGTGTGCTAAT 1379
Qy 418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db 1380 CGAGTCAATCCAAAGAGTGAATGTAAGTCTGTCTACAAAACCCACATTTGATGCATT 1439
Qy 438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db 1440 CATTATCGGAACAGATGCAAAACGTCTGCAT----- 1472
Qy 458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db 1473 -----GGCCTTTGATGAAGAGCAGAACAGAAACTTTTTCAGAGAAA 1514
Qy 478 LysIleGlnLysLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db 1515 CGTGTGGAATTGCTTAAGGAACCTTCAGGAAACCCAGACATTTATGAGAGGCTTGCTCA 1574
Qy 498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db 1575 GCCTTGCTCCAGCATTTATGAACATGAAGATATAAAGAGGAATTTGCTTCAGCTC 1634
Qy 518 PheGlyGlyLysAlaLysLysIle-----ProSerGlyAlaSerPheArgGlyAspIle 535
Db 1635 TTTGGCGGACAGGAAGGATTTTAGTCACATCGGAAGGGCAAAATTCGGCTGAGATC 1694
Qy 536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuGlnTyrValHis 555
Db 1695 AACATCTTGCTGTGTGGCGACCTTGCTACAGCAAGTCCAGCTGCTGCACTAGTACGTGAC 1754
Qy 556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db 1755 AACCTCGTCCCAAGGGCCAGTAGTACAGCTCTGGGAAGGGCTCCAGTGGCTGCTCACT 1814
Qy 576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
```

RESULT 11

ABL65666

ID ABL65666 standard; DNA; 3273 BP.

XX AC ABL65666;

XX AC ABL65666;

DT 15-MAY-2002 (first entry)

XX Lung cancer related gene sequence SEQ ID NO:4003.
DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; da.
XX Homo sapiens.
XX WO200194629-A2.
XX 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US010838.
XX 05-JUN-2000; 2000US-0209473P.
XX 05-JUN-2000; 2000US-0209531P.
XX 18-SEP-2000; 2000US-0231313P.
XX 18-SEP-2000; 2000US-0233617P.
XX 20-SEP-2000; 2000US-0234009P.
XX 20-SEP-2000; 2000US-0234034P.
XX 20-SEP-2000; 2000US-0234052P.
XX 22-SEP-2000; 2000US-0234509P.
XX 22-SEP-2000; 2000US-0234567P.
XX 25-SEP-2000; 2000US-0234923P.
XX 25-SEP-2000; 2000US-0234924P.
XX 25-SEP-2000; 2000US-0235077P.
XX 25-SEP-2000; 2000US-0235082P.
XX 25-SEP-2000; 2000US-0235134P.
XX 25-SEP-2000; 2000US-0235280P.
XX 26-SEP-2000; 2000US-0235637P.
XX 26-SEP-2000; 2000US-0235638P.
XX 27-SEP-2000; 2000US-0235711P.
XX 27-SEP-2000; 2000US-0235720P.
XX 27-SEP-2000; 2000US-0235840P.
XX 27-SEP-2000; 2000US-0235863P.
XX 28-SEP-2000; 2000US-0236028P.
XX 28-SEP-2000; 2000US-0236032P.
XX 28-SEP-2000; 2000US-0236033P.
XX 28-SEP-2000; 2000US-0236034P.
XX 28-SEP-2000; 2000US-0236109P.
XX 28-SEP-2000; 2000US-0236111P.
XX 29-SEP-2000; 2000US-0236842P.
XX 29-SEP-2000; 2000US-0236891P.
XX 02-OCT-2000; 2000US-0237172P.
XX 02-OCT-2000; 2000US-0237173P.
XX 02-OCT-2000; 2000US-0237278P.
XX 02-OCT-2000; 2000US-0237294P.
XX 02-OCT-2000; 2000US-0237295P.
XX 03-OCT-2000; 2000US-0237316P.
XX 03-OCT-2000; 2000US-0237425P.
XX 03-OCT-2000; 2000US-0237598P.
XX 03-OCT-2000; 2000US-0237604P.
XX 03-OCT-2000; 2000US-0237606P.
XX 03-OCT-2000; 2000US-0237608P.
XX 01-NOV-2000; 2000US-0244867P.
XX 01-NOV-2000; 2000US-0245084P.
XX (AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX Claim 1; SEQ ID NO 4003; 44pp; English.
XX The present invention describes a method (M1) for screening for an anti-

CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 6,4e-98 Length: 3273
Score: 1817.50 Matches: 406
Percent Similarity: 57.78% Conservative: 151
Best Local Similarity: 42.12% Mismatches: 256
Query Match: 39.43% Indels: 151
Gaps: 6
US-10-768-511-8 (1-901) x ABL65666 (1-3273)
QY 38 AlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArg 57
Db 21 GCAGGTCGTCGCAAGCGCGCGC-----CTTCCAGCTAACGCG---CGC 62
QY 58 ArgArgGlyArgGluThrAspSerAlaArgArgArgSerArg----- 72
Db 63 CGCGCGGCGAGGCGGTGGCGCGGAGCGCGGAAAGCTCCGCGCTCGCGAGCAGGCGAG 122
QY 73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db 123 GGAAGCGCGGAGGCGGCG---CCGCGCGCGAGCTTGTCTTCTCGCGCAGGACTCCGAGC 179
QY 92 ThrProGlyThrProValAlaThrProVal-----TyrAlaThr 104
Db 180 ACTATGTCGTCCTCCGCGCTCGAGCCCGGAGCGCGCGCGCGCGCTGGAGGCGCCACC 239
QY 105 ProValGlyThrPro----- 109
Db 240 CCGCGCCAGAGCGCTCGGAGTGAGGATGCCAGGTCACTCTCCCTCTCAGAGAGTAGAGGC 299
QY 109 ----- 109
Db 300 GAGGATTCACCTCCACGCGGAGTTGCAGCGGATGCCAACCTCGCCTGGAGTGGACCTG 359
QY 109 ----- 109
Db 360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTCCAGCCCTCCCCAAATGCAATTTTCAGCT 419
QY 110 -----MetGlyThrProSerPheHis--- 116
Db 420 ATCCCTCTTGACTTTGATGTTAGTTTACCACCTGACATACGCGCACTCCCGAGCTCTCGGTA 479
QY 117 -----ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db 480 GAGGGAACCCCAAGAAGTGGTGTAGGGGCGACACT---GTGAGACAGAGCGCTGACCTG 536
QY 129 GlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluProGlyHis 148
Db 537 GGCTCTGCACAGAAGGCGCTGCAAGTGATCTGCAGTCT----- 575
QY 149 ArgSerProSerArgGluProSerAlaAspGlyArgProGluSerAlaGluProAsp 168
Db 576 -----GACGGGCGCAGCAGCAGAGATATAGTGGCAAGT 608

Db 2694 GATATGTTTGAAGAAGCACTGGTGCCTCGCAGATGATGATTTCTTGACAGTGACTGGG 2753

Qy 896 AepIleVallye 899

Db 2754 AAGACCGTGC GC 2765

RESULT 12

ABL65259

ID ABL65259 standard; DNA; 3273 BP.

XX

AC ABL65259;

XX

XX 15-MAY-2002 (first entry)

XX

DE Lung cancer related gene sequence SEQ ID NO:3596.

XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

KW gene; ds.

XX

OS Homo sapiens.

OS

PN WO200194629-A2.

XX

PD 13-DEC-2001.

XX

PF 30-MAY-2001; 2001WO-US010838.

XX

PR 05-JUN-2000; 2000US-0209473P.

PR 05-JUN-2000; 2000US-0209531P.

PR 18-SEP-2000; 2000US-0233133P.

PR 18-SEP-2000; 2000US-0233133P.

PR 20-SEP-2000; 2000US-0234009P.

PR 20-SEP-2000; 2000US-0234034P.

PR 20-SEP-2000; 2000US-0234052P.

PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.

PR 25-SEP-2000; 2000US-0234923P.

PR 25-SEP-2000; 2000US-0234924P.

PR 25-SEP-2000; 2000US-0235077P.

PR 25-SEP-2000; 2000US-0235082P.

PR 25-SEP-2000; 2000US-0235134P.

PR 25-SEP-2000; 2000US-0235280P.

PR 26-SEP-2000; 2000US-0235637P.

PR 26-SEP-2000; 2000US-0235638P.

PR 27-SEP-2000; 2000US-0235711P.

PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.

PR 27-SEP-2000; 2000US-0235863P.

PR 28-SEP-2000; 2000US-0236028P.

PR 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.

PR 28-SEP-2000; 2000US-0236034P.

PR 28-SEP-2000; 2000US-0236109P.

PR 28-SEP-2000; 2000US-0236111P.

PR 29-SEP-2000; 2000US-0236842P.

PR 29-SEP-2000; 2000US-0236891P.

PR 02-OCT-2000; 2000US-0237172P.

PR 02-OCT-2000; 2000US-0237173P.

PR 02-OCT-2000; 2000US-0237278P.

PR 02-OCT-2000; 2000US-0237294P.

PR 02-OCT-2000; 2000US-0237295P.

PR 02-OCT-2000; 2000US-0237316P.

PR 03-OCT-2000; 2000US-0237425P.

PR 03-OCT-2000; 2000US-0237598P.

PR 03-OCT-2000; 2000US-0237604P.

PR 03-OCT-2000; 2000US-0237606P.

PR 03-OCT-2000; 2000US-0237608P.

PR 01-NOV-2000; 2000US-0244867P.

PR 01-NOV-2000; 2000US-0245084P.

PA (AVAL-) AVALON PHARM.

XX

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;

PI Soppet DR, Weaver Z;

XX

DR WPI; 2002-188264/24.

XX

XX Screening for anti-neoplastic agent involves exposing cells to a chemical

PT agent to be tested for anti-neoplastic activity, and determining a change

PT in expression of a gene of a signature gene set.

XX

XX Claim 1; SEQ ID NO 3596; 44pp; English.

XX

CC The present invention describes a method (M1) for screening for an anti-

CC neoplastic agent. The method involves exposing cells to a chemical agent

CC to be tested for anti-neoplastic activity, determining a change in

CC expression of at least one gene (I) of a signature gene set, where (I)

CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664

CC to ABL70110), or is at least 95% identical to (S), where a change in

CC expression is indicative of anti-neoplastic activity. (I) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening an

CC anti-neoplastic agent, and can be used for producing a product which is

CC the data collected with respect to the anti-neoplastic agent as a result

CC of M1, and the data is sufficient to convey the chemical structure and/or

CC properties of the agent. M1 can be used in the treatment of cancer such

CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,

CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell

CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous

CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's

CC tumour

XX

SQ Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

XX

Alignment Scores:

Pred. No.: 6.4e-98 Length: 3273

Score: 1817.50 Matches: 406

Percent Similarity: 57.78% Conservative: 151

Best Local Similarity: 42.12% Mismatches: 256

Query Match: 39.43% Indels: 151

DB: 6 Gaps: 18

US-10-768-511-8 (1-901) x ABL65259 (1-3273)

Qy 38 AlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArg 57

Db 21 GCGAGCGTGTCCGCAAGCGCCGC-----CTTCCACGTAAACGCG---CGC 62

Qy 58 ArgArgGlyArgGluThrAspSerAlaArgArgArgSerArg----- 72

Db 63 CGGCGGGGAGGGCGTTCGCGCGGAGCGCGGACGTCGCGTTCGGAGCAGGGCAG 122

Qy 73 ---SerArgSerLeuGlyAsnSerValTySerSerProTyArgAlaGlyThrProGly 91

Db 123 GGAAGCCGGAGGCGGGC---CCGCGCCGAGCTTGTCTTTCGCGCAGGTACTCCGAGC 179

Qy 92 ThrProGlyThrProValAlaThrProVal-----TyxAlaThr 104

Db 180 ACTATGTCTCCCGCGGTTCGACCCCGAGCCGCGCGCAGCGCGCGTGTGAAGGGCCACC 239

Qy 105 ProValGlyThrPro----- 109

Db 240 CCGCGCCAGACGCTTCGGAGTGAGATGCCAGGTCACTCCCTCTCAGACGTAGAGGC 299

Qy 109 ----- 109

Db 300 GAGGATTCACCTCCACGGGGAGTTCAGCGCGATGCCAACCTCGCTGGAGTGGACCTG 359

Qy 109 ----- 109

Db 360 CAGAGCACTGTCTGCGCAGGACGTGCTGTTTCCAGCCCTCCCCAAATGATTTTCAGCT 419

Qy 110 -----MetGlyThrProSerPheHis--- 116

Db 420 ATCCCTCTTGACTTTGATTTAGTTTACCACTGACATACGGCACTCCCGGCTCTCGGTA 479

QY 117 -----ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db 480 GAGGAACCCAGAGAGTGGTGTAGGGGCACACT---GTGAGACAGAGCGCTGACCTG 536
QY 129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db 537 GGCCTCTGCACAGAGGGCTGCAAGTGGATCTGCAGTCT----- 575
QY 149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db 576 -----GACGGGGCAGCAGCAGAGATATAGTGCACAGT 608
QY 169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db 609 GAGAGCTCTCTAGGGCCAAAACCTGTGATCTGGGAACAGATGTAATGTGGCAGCATGC 668
QY 189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----Arg 200
Db 669 AAAGAAACTTTCAGAGATTTCTTCAGCGTATTATTGACCTCTGGCTAAAGAGAGAA 728
QY 201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db 729 AATGTGGCATAGATATTACTGAACCTCTATACATCAACGACTTGGGGAGATTAATGTT 788
QY 220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db 789 ATTGGTGAGCAATTTTAAATGTGAACCTGTGAACACATCAATCAATTTGACAAAATTTG 848
QY 240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db 849 TACAGACAACTCATCTCTTACCACAGGAAGTTATTCCAACCTTTTGACATGGCTGTCAAT 908
QY 260 GluValAlaThrSerLeuLeuPro-----ThrPheGluLysHisIleGluAlaArgPro 277
Db 909 GAATCTCTTTGACCGTTACCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968
QY 278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db 969 TTCACGGCATTGAAGACTAAGATATAGAAACCTGAATCCAGAAACATGACACAGTCC 1028
QY 298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db 1029 ATCACCATCAGCGCATGGTGTATCAGGACATCCAGCTGATTCCTCCAGATGACGAGGCC 1088
QY 318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValLysGlyArg 337
Db 1089 TTCTTCAGTGGCCAGTGTGTGCCACAGC---ACCGGGGTGAGATGACCGCGCCGC 1145
QY 338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db 1146 ATTGACAGCCAGTGTGTGGGGCGC-----TGCCACACCCACACGATGGCAGTCTC 1199
QY 358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db 1200 ATCCACACCGCTCCCTTCTCTGACAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259
QY 378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db 1260 GACATGCTGTCAGGGCAGACACACACACACATGTTATCTGTTGCTCACAATGATCTCGTT 1319
QY 398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db 1320 GACAGGTCCAGCTGGGGACAGATGTAATGTTACAGGCAATCTATCAGGCTGTGCCTATT 1379
QY 418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db 1380 CGAGTCAATCAAGAGTGAATGTAAGTGTCTACAAAACCCACATGATGATTCATT 1439
QY 438 HisValLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db 1440 CATTATCGGAAACGGATGCAAAAGCTTCGAT----- 1472

QY 458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db 1473 -----GGCCTTGTATGAAGACAGACAGAACTTTTTTTCAGAGAA 1514
QY 478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db 1515 CGTGTGAATTCCTTAAGGAACCTTCCAGGAACACAGACATTTATGAGAGCTTGTCTCA 1574
QY 498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLysCysGlnLeu 517
Db 1575 GCCTTGGCTCCAGCATTTATGAACATATAAAGAGGAAATTTTGTCTCAGCTC 1634
QY 518 PheGlyGlyLysAlaLysLysIle-----ProSerGlyAlaSerPheArgGlyAspIle 535
Db 1635 TTTGCGGGCAAGAGAGATTTTAGTCACACTGGAAGGGCAAAATTTTCGGCTCAGATC 1694
QY 536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGluTyrValHis 555
Db 1695 AACATCTTGTGTGTGGCGACCTCGTACCAAGTCCCGAGTCTCGAGTACGTGTAC 1754
QY 556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db 1755 AACCTGCTCCCGAGGGCCAGTACACGCTCGGAAGGGCTCCAGTGCAGTGGCTCACT 1814
QY 576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db 1815 GCGTACGTAAATGAAGACCTCGAGACAAGCAGCTGGTCTCTGCACAGGTGCTCTGTGC 1874
QY 596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db 1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAGTCAAGA 1934
QY 616 SerMetLeuHisGluValMetGluGlnThrValSerValAlaLysGlyLysIleIle 635
Db 1935 TCGGTATTGTCATGAAGTATGGAACAGCAGACTCTGTCCATTGCAAGAGCTCGGATCATC 1994
QY 636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db 1995 TGTCAGCTCAATGCGCGCACCTCTGTCTGGCAGCAGCAATCCCATTCAGTCTCAGTGG 2054
QY 656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThrLeuLeuSerArgPhe 675
Db 2055 AATCTCTAAAACCAACCATTTGAACATCCAGCTCGCTCATATTTATTATCAAGTTT 2114
QY 676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db 2115 GATTTGATCTTCTCATGCTGACCTCAGGACGAGGCTATGACAGGCGCTCTGGCTCAC 2174
QY 696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db 2175 CACCTGGTGCATCTGTACTACAGAGGAGGAGGAGGAGGAGGAGTCTCTGGACATG 2234
QY 716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db 2235 GCGGTGCTAAAGCATACATTCCTACGCGCAGCAGCACCATCATCGCGGCTAAGTGAG 2294
QY 736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db 2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----ATT 2342
QY 756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db 2343 GGCAGTAGCGGGAATGTTTCTGCATACCTCCGACAGTACAGTATTAATCCGCTTA 2402
QY 776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db 2403 GCAGAGACCCATGCTAAAGTATGTTTAAACAAAGTTGAAGCCATCATGATGGAGAG 2462
QY 796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db 2463 GCCAAACGGCTTCATCGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGAGTGGCATC 2522
QY 816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835


```
QY 358 IleHisAsnArgCysThrPheAlaHisLysGlnIleValArgLeuGlnGluThrProAsp 377
DB 1200 ATCCACACCGCTCCCTCTCTCTGACAGCAGATGATCAAGCTTCAGGAGTCTCGGAA 1259
QY 378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuThrMetVal 397
DB 1260 GACATGCTCCGAGGCGACACACACACAGTATATCTGTGTGCTCACAATGATCTCGTT 1319
QY 398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
DB 1320 GACAGGTCACGCTCGGAGCAGAGTGAATGTATACAGGATCTATCGAGCTGTGCTATT 1379
QY 418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
DB 1380 CGAGTCAATCCAAGAGTGAGTGAAGTCTGTCTACAAACCCACATTCATGTCATT 1439
QY 438 HisValLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
DB 1440 CATTATCGGAAACGAGTCAAAACGCTCTGCAT----- 1472
QY 458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaHisGluAla 477
DB 1473 -----GCGCTTGATGAAGAGCAGACAGAACTTTTTCAGAGAAA 1514
QY 478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
DB 1515 CTTGTGGAAATGCTTAAGAACTTTCAGAAACCCAGACATTTATGAGAGGCTTGCTTCA 1574
QY 498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
DB 1575 GCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAGGAATTTTGTCTCAGCTC 1634
QY 518 PheGlyGlyLysAlaLysLysIle-----ProSerGlyAlaSerPheArgGlyAspIle 535
DB 1635 TTTGGCGGACAGGAAGAGTATTTAGTCACACTGGAAGGGGCAAAATTTCCGGCTCAGATC 1694
QY 536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
DB 1695 AACATCTTCTGTGTGGCCACCTTGATCCAGAACGTCCTGCTGTCAGTACGATGATAC 1754
QY 556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
DB 1755 AACCTGCTCCAGGGCCAGTACAGCTCTGGAGGGCTCCAGTGCAGTGGCTCACT 1814
QY 576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
DB 1815 GGTACGTAATGAAGACCTCGAGCAAGCAGCTGCTCTGCAGCAGCTGCTCTGTC 1874
QY 596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
DB 1875 CTGAGTGACACGGCATCTGCTGTATCGATGATTCGACAGATGAATGAAGATCACA 1934
QY 616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
DB 1935 TCGTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1994
QY 636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
DB 1995 TGTGAGTCAATGCGCGCACCTCTGCTGCGCAGCAGCAAAATCCCATTTGAGTCTCAGTGG 2054
QY 656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
DB 2055 AATCCTAAAAAACAACCATTTGAACATCCAGCTGCCTCATATCTTTATTATCAAGGTTT 2114
QY 676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgGlyLeuAlaArg 695
DB 2115 GATTTCATCTCTCTGATGAGTGGACCTTCAGGAGCAAGCCCTATGACAGGCGCTCGGCTCAC 2174
QY 696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
DB 2175 CACCTGGTGCATCTGATACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2234
QY 716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
```

```
DB 2235 GCGGTGCTAAAGGAGCTACATTGCTTACGCGCACAGCACCATCATCCGCGGCTAAGTGAG 2294
QY 736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
DB 2295 GAAGCCACCCAGGCTCTCATCGAGCTTATGTAGACATGAGNAG-----ATT 2342
QY 756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
DB 2343 GGCAGTAGCCGGGAATGTTCTTCATACCTCCGACAGCTAGAGTCATTATATCCGCTTA 2402
QY 776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
DB 2403 GCAGAAGCCCATGCTTAAAGTAAAGTATGCTTACAAAGTTGAAGCCATGATGTGAAGAG 2462
QY 796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
DB 2463 GCCAAACGCTTCATCGGGAAGCTCTGAAGCAGTCTGCACACTGATCCCGGACTGGCATC 2522
QY 816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgAlaAsn 835
DB 2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCTGTAACGGAAGAGAA 2582
QY 836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
DB 2583 TTAGCTGAAGCATTCGAAAAGCTTATT-----TTATCTAAGGGCAAAACACAGCT 2633
QY 856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
DB 2634 CTAAATAATCCAGCAACTTTTGAAGATATTCGGGACAAATCTGCATAGCAATTAATAA 2693
QY 876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
DB 2694 GATATGTTGAAGAGCACTGCTGCCCTGGCAGATGATGATTTCTGCAGTACTGGG 2753
QY 896 AspIleValLys 899
DB 2754 AAGACCGTGGC 2765
DB 2754 AAGACCGTGGC 2765
RESULT 15
ADR26059
ID ADR26059 standard; DNA; 3273 BP.
XX ADR26059;
AC ADR26059;
XX 21-OCT-2004 (first entry)
XX Breast cancer prognosis marker #1920.
XX ds; breast cancer; prognosis; gene expression; diagnosis.
XX Homo sapiens.
XX WO2004065545-A2.
XX 05-AUG-2004.
XX 15-JAN-2004; 2004WO-US001100.
XX 15-JAN-2003; 2003US-00342887.
XX (ROSE-) ROSETTA INPHARMATICS LLC.
XX (NECA-) NETHERLANDS CANCER INST.
XX Van't Veer LJ, He Y;
XX WPI; 2004-593473/57.
XX
XX Classifying a breast cancer patient according to prognosis comprises
XX determining the similarity between the level of expression of each of
XX five genes in a cell sample taken from patient, to control levels.
XX Disclosure; SEQ ID NO 1920; 226pp; English.
```

XX The invention relates to a method of classifying a breast cancer patient
 CC according to prognosis by determining the similarity between the level of
 CC expression of each of five genes for which markers are listed in the
 CC specification, in a cell sample taken from the breast cancer patient, to
 CC control levels of expression for each respective five genes to obtain a
 CC patient similarity value. The methods are useful for classifying a breast
 CC cancer patient according to prognosis. Kits and computer program products
 CC are useful for data analysis using the diagnostic, prognostic and
 CC statistical methods of the invention. This sequence corresponds to a
 CC marker used in the method of the invention.

XX Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.4e-98	Length:	3273
Score:	1817.50	Matches:	406
Percent Similarity:	57.78%	Conservative:	151
Best Local Similarity:	42.12%	Mismatches:	256
Query Match:	39.43%	Indels:	151
DB:	13	Gaps:	18

US-10-768-511-8 (1-901) x ADR26059 (1-3273)

```

Qy 38 AlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArg 57
Db 21 CGCAGCTGTCGGCAAGCGCGCGC-----CTTCCACGTAACGCG---CGC 62

Qy 58 ArgArgGlyArgGluThrAspSerAlaArgArgArgSerArg-----72
Db 63 CGCGGGGGAGGCGCTTGGCGGAGCGAGCGAGCGGAACTCCGCGTGGGAGCAGGCGAG 122

Qy 73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db 123 GGAAGCCGGAGGCGGCG---CCGCGCCGAGCTTCTCTTGTGCGCGAGGTACTCCGAGC 179

Qy 92 ThrProGlyThrProValAlaThrProVal-----TyrAlaThr 104
Db 180 ACTATGTCGTCCCGCGCTCGACCCCGAGCGCGCGCGCGCGCGCTGGAGGCGCCACC 239

Qy 105 ProValGlyThrPro-----109
Db 240 CCGCGCCAGACGCTCGGAGTGAGGATGCCAGGTCACTCCCTCTCAGAGAGGTAGAGGC 299

Qy 109 -----109
Db 300 GAGGATTCCACCTCCACGGGGAGTTGCAGCGCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy 109 -----109
Db 360 CAGAGCACTGTCGCGAGGACGTGCTGTTTTCAGCCCTCCCAATGCACTTCTTCAGCT 419

Qy 110 -----MetGlyThrProSerPheHis---116
Db 420 ATCCCTCTTGACTTTGATGTTAGTTTCCACCTGACATACGGGACCTCCAGCTCTCGGTA 479

Qy 117 -----ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db 480 GAGGGAAACCCCAAGAAGTGTGTAGGGGCACACCT---GTGAGACAGAGGCGCTGACCTG 536

Qy 129 GlySerGlnGlyProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db 537 GGCCTCTGCACAGAGGCGCTGCAAGTGGATCTGCAGTCT-----575

Qy 149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db 576 -----GACGGGGGAGCAGCAGAGAGATATAGTGGCAAGT 608

Qy 169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db 609 GAGCAGTCTCTAGGCCAAACAACTTGTGATCTGGGGAACAGATGTAATGTTGGCAGCATGC 668

Qy 189 LeuArgAlaIleArgPheLeuHisAsnTyr-----Arg 200

```

```

Db 669 AAAGAAAACCTTTCAGAGATTCTTCAGCGTTTTATGACCTCTGGCTAAAGAGAAGAA 728
Qy 201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db 729 AATGTTGGCATAGATATTACTGACCTCTATACATCAACGACTTGGGGAGTAATGTT 788
Qy 220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db 789 ATTGGTGAGCAATTTTAAATGTGAACATGTGAACACATCAATCATTTTGACAAAATTTG 848
Qy 240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db 849 TACAGACAATCATCTCTTACCCACAGAGATTTATCCAACTTTTGATCGGTGCTCAAT 908
Qy 260 GluValAlaThrSerLeuLeuPro-----ThrPheGluLysHisIleGluAlaArgPro 277
Db 909 GAAATCTTCTTGGACCGTTACCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968
Qy 278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db 969 TTCAACGCATTGAAGACTAAGATATGAGAAACCTGAATCCAGAAGACATTTGACACGCTC 1028
Qy 298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db 1029 ATCACCATCAGCGCATGGTGATCAGGACATCCCGAGTATCCCGAGATCGAGGAGCC 1088
Qy 318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db 1089 TTCTTCCAGTGCCAAAGTGTGTGCCACACG---ACCCGGGTGGAGATGACCGCGCGCGC 1145
Qy 338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db 1146 ATTGACAGGCGCAGTGTGTGGCGCGC-----TGCCACACACACACACATGTCACCTC 1199
Qy 358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db 1200 ATCCACACCGCTCCCTCTCTGACAGCAGATGATCAAGCTTTCAGAGGTCTCCGAAA 1259
Qy 378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db 1260 GACATGCTGCGAGGCGCAGACACACACACTTATCTGTTTGTCTCAATGATCTCGTT 1319
Qy 398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db 1320 GACAAGTTCAGCTGCGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCTATT 1379
Qy 418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db 1380 CGAGTCAATCCAGAGTGAAGTAATGTGAAGTCTGTCTACAAAACCCACATGATGTCATT 1439
Qy 438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db 1440 CATATCCGAAAACGGATGCAAAACGCTGCAAT-----1472
Qy 458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db 1473 -----GGCCTTGATGAAGAAGCAGAACAGAACTTTTTCAGAGAAA 1514
Qy 478 LysIleGlnLysLeuLysGluSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db 1515 CGTGTGAATGCTTAAGGAACTTTCCAGGAAACACAGATTTATGAGAGGCTTCTCTCA 1574
Qy 498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db 1575 GCCTTGGCTCCAGCATTTATGAACATGAAGATATAGAAGAGGAATTTTGTCTCAGCTC 1634
Qy 518 PheGlyGlyLysAlaLysLysIle-----ProSerGlyAlaSerPheArgGlyAspIle 535
Db 1635 TTTGCGGGAACAAGAGAGGATTTTAGTCACTCGAAGGGGCAAAATTTTCGGGCTCGAGATC 1694
Qy 536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555

```

Search completed: December 6, 2005, 10:15:30
Job time : 972 secs

```
Db 1695 AACATCTTGTGTGTGGCGACCTGGTATACAGCAAGTCCAGCTGCTGTCAGGTACGTGTAC 1754
Qy LysIleAlaProArgGlyIleThrSerGlyArgGlySerSerAlaValAlaGlyLeuThr 575
Db 1755 AACCTGTGCCCCAGGGCCAGTACAGCTGTGGAAAGGGCTCCAGTGCAGTTGGCCCTCACT 1814
Qy 576 AlatyValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db 1815 GCGTAGCTAATGAAGACCTCGAGACAAGGCGAGTGGTCTCGACAGCAGGTGCTCTGTGC 1874
Qy 596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db 1875 CTGAGTGCACACGGCATCTGCTGTATCGATGATGTTCCACAGATGAATGAAGATCAAGA 1934
Qy 616 SerMetLeuHisGluValMetGluGlnThrValSerValAlaLysGlyIleIle 635
Db 1935 TCGGTATTGCATGAAGTATGGAACAGCAGACTCTGTCATTGCAAGGCTGGGATCATC 1994
Qy 636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db 1995 TGTCACTCAATGGCGCACCTCTGTCTGGCAGCAGCAAAATCCCATTTGAGTCTCAGTGG 2054
Qy 656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db 2055 AATCTTAAAAAACAACCATTCGAAACATCCAGCTGCTCTACATCTTTATTATCAAGGTTT 2114
Qy 676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db 2115 GATTTGATCTTCTCATGCTGACCTCAGAGCAAGCCTATGACAGCGGCTCTGGCTCAC 2174
Qy 696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db 2175 CACCTGTGCGACTGTACTACAGAGCAGGAGGAGCAGGAGGAGGAGCTCTGGACATG 2234
Qy 716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db 2235 GCGGTGCTAAGGACTACATTCGTACGCGGCACAGCACCATCATGCGCGGCTTAAGTGAG 2294
Qy 736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db 2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----ATT 2342
Qy 756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db 2343 GGCAGTAGCGGGGAATGTTCTGTATACCTTCGACAGCTAGAGTCATTATCCGCTTA 2402
Qy 776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db 2403 GCAGAAGCCCATGCTAAAGTAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462
Qy 796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db 2463 GCCAAACGGCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCGGACTGGCATC 2522
Qy 816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db 2523 GTGGACATATCTATTCTTACGGGGATGAGTGCCACCTCTCTGTAACGGAAAGAGAA 2582
Qy 836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db 2583 TTAGCTGAAGCATTTGAAAGAGCTTATT-----TTATCTAAGGGCAAAACACCCAGCT 2633
Qy 856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db 2634 CTAATAATACCGCAACTTTTGAAGATATTTCGGGGACAATCTGACATAGCAATTACTAAA 2693
Qy 876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db 2694 GATATGTTGAAGAGCAGCTGCGCTGGCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753
Qy 896 AspIleValLys 899
Db 2754 AAGACCGTGGC 2765
```

THIS PAGE BLANK (USPTO)


```

Db 257 RGENTRLAQPGDHVIGITVFLPM-LRTGFRQLVQGLLSETYLECHRLVKMNKSEDELGT 315
QY 450 EDPMEMDKENDMYAGYHESDTSAAANEAKIKELSKLPGIYDRLSRLAPSWELEDI 509
Db 316 BELSE-----BELROITE-EDFYEKLAASIAPFIYGHEDV 349
QY 510 KGLLCQLFGGKAKKIPGASFRGDIINVLVDPGTSQQLQYVHKIAPRGIYTSGRGS 569
Db 350 KKALLLLLVGG-VDNSPRGMKIRGINICLMGDPGVAKSQQLSYIDRLAPRSQYTTGRGS 408
QY 570 SAVGLTAYVTKDPETRETIVLESALVLSDRGICCCIDEFKMSDNARSMLHEVMEQOTVS 629
Db 409 SGVGLTAAMVDPVVTGEMTLEGALVLADQGVCCIDEFDKMDTDTRTAIHEVMEQOTISI 468
QY 630 AKGGIIASINARTSVLACANPSGSRNARLSVIDNIQLPPTLLSRFDLIYMLDKPDRON 689
Db 469 AKAGIMTILNARCSILAAANPAYGRNPKTVQEQNIQPAALLSRFDVLWLIQDKPDRDN 528
QY 690 DRLARHLVALHYENYVSKQ-DALDLOTLTAYITAYARQHVHTLSDBAAEDLINGVEM 748
Db 529 DLRLAQHITYVHOHSKQPPSQPQLMDKLMRRYITMCKRK-QPAIPEALADYLTAAYVEM 587
QY 749 RQKGNFPSSKKVITATPROLESMTIRSEALARMFSEVVEKVDAAEAVRLDLVALQOSA 808
Db 588 RKEAR---TNKDMTFTSARTLLSVLRSTALARLEVDVEKEDVNEAMRLMEMSKDSLL 644
QY 809 TDHATGTIDMDLITTVGSASERIRANLL-LALRELIADK 847
Db 645 GDKG-----HTSRTORPADVIFSTIREWVPEK 671

RESULT 15
JC4580
replication licensing factor MCM7 - mouse
N:Alternate names: cell division control protein CDC47
C:Species: Mus musculus (house mouse)
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: JC4580
R:Takizawa, N.; Kimura, H.; Sugimoto, K.
Gene 167, 343-344, 1995
A:Title: Sequence of mouse CDC47, a member of the minichromosome maintenance (Mcm) family
A:Reference number: JC4580; MUID:96144303; PMID:8566808
A:Accession: JC4580
A:Molecule type: mRNA
A:Residues: 1-719 <TAK>
A:Cross-references: UNIPROT:Q61881; UNIPARC:UPI0000029103; DDBJ:D26091; NID:g1136746; PI
C:Comment: The complex of six MCM proteins is one of several proteins that must be bound
phosphorylated and dissociate from the chromatin.
C:Genetics:
A:Gene: cdc47
C:Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C:Function:
A:Description: part of the replication licensing system that permits DNA replication to
C:Superfamily: replication licensing factor MCM7; MCM homology
C:Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F:149-636/Domain: MCM homology <MCM>

Query Match 21.0%; Score 966; DB 1; Length 719;
Best Local Similarity 32.6%; Pred. No. 1.6e-46;
Matches 236; Conservative 132; Mismatches 221; Indels 134; Gaps 19;

QY 192 IRRFLHN--YRSSADHLNSKIYIETVEEREDTLNIDMSDIYDHPDPLVAKIVRYPLD 249
Db 13 VKKLFQEFYENELGKKQFKYGTQVLVHLAHRQVALYVDLDDIAEDDFE----- 61
QY 250 IIPLDTECQ-----EVATSLPTFEK-----HIEAR----- 276
Db 62 ---LVDSICENAKRYSLRFGDVVQELLPEYKEVKNVDLVYIEHRLMMEQRSDPCA 118
QY 277 -----PFNLKASVHMRNLNPSDIDKLVSVKGMVTRCSSIIPEI 314
Db 119 VRNPQNPSELNRRRPFELYFRGSSSKPRV-IREVRADSVGKLLTVRGIVTRVSEVKPRM 177

```

```

QY 315 KGAPFKCLVCG---HSPPLVTWVKGRVBEPT-----RCEKPECAARNA---MSLIHNRCT 363
Db 178 VVAITYTCDQCAGTYQP-----IQSPFIMPLIMCPSQECQTNRSRGGRLYLQTRGSK 228
QY 364 FANKQIVRLQOTTPDAIPEGETPHVTSMCLYNTWVDVAVKPGDRIIBVTGVFKAMAVRVGNQ 423
Db 229 FVKFQEMKIQBSDOQVPVGNIPRSITVVLEGENTRIAPQPGDHVSVTGIFLPV-LRTGFGQ 287
QY 424 RTLAALYKTYIDCVHVVKSDRGLQTEDEPMEMDKENDMYAGYHESDTSAAANEAKIKELK 483
Db 288 MAQGLLSETYLEAHWIVK-----MTKSDDDVSG-----AGELSSELK 325
QY 484 ELSKLPGIYDRLSRLAPSWELEDIKKGLLQCLFGGKAKKIPGASFRGDIINVLVDPG 543
Db 326 QIAB-EDFYEKLAASIAPFIYGHEDVKKALLLVGG-VDQSPQGMKIRGINIICLMGDP 383
QY 544 GTSKQLQYVHKIAPRGIYTSGRSSAVGLTAYVTKDPETRETIVLESALVLSDRGICCC 603
Db 384 GVAKSQQLSYIDRLAPRSQYTTGRSGVGLTAAVLRDSVSGELTLEGALVLADQGVCC 443
QY 604 IDEFDKMSDNARSMLHEVMEQOTVSVAKGIIASINARTSVLACANPSGSRNARLSVID 663
Db 444 IDEFDKMSADRTAIHEVMEQOTISIAGAGITTLNARCSILAAANPAYGRNPFARSLEQ 503
QY 664 NIQLPPTLLSRFDLIYMLDKPDEQNDRRLARHLVALHYENYVSKQ-DALDLOTLTAYI 722
Db 504 NVQLPAALLSRFDLLWLIIQDRPDRNDLRLAQHITYVHOHSKQPPAQPEPLDKLMRRYI 563
QY 723 TYARQHVHTLSDBAAEDLINGVEMRQKGNFPSSKKVITATPROLESMTIRSEALARM 782
Db 564 AMCHER-OPTVPESLADYITAAAYVEMRREAR---ASKDATTYSARTLLAILRLSTALARL 619
QY 783 RFSEVVEKVDAAEAVRLDLVALQOSAETHATGTIDMDLITTVGSASERIRANLLAALRE 842
Db 620 RMVDIVEKEDVNEAIRLWEMSK-----DSSLGKGGQTARTORPADVIFATIRE 667
QY 843 LIA 845
Db 668 LVS 670

```

Search completed: December 7, 2005, 17:43:43
Job time : 49 secs

Db 722 -VGAGRGQISAYPRQLESILRLSEAHAKVRLSNQVELLDVEARHLREALQSDTDLPLS 780
QY 814 GTYDMDLITTVGASERIRANLLAALRELIADKISPGSSGLKTSQLEDTRSSQSDV 873
Db 781 GKIDVGILITGLSTAARKKRALVAIKENLKKK---GKVLTVPYQKLFSDIKEGSIQMI 837
QY 874 SLQDIKNALGSLQGGFLTVHGDIVKRV 901
Db 838 TREQFEDALKEVQDEGAIVVMGKNVTIRI 865

RESULT 9
T50408
cdc21 protein [imported] - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 02-Sep-2000
C:Accession: T50408
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, February 2000
A:Reference number: Z25039
A:Accession: T50408
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-407 <SEE>
A:Cross-references: UNIPARC:UPI00001690AB; EMBL:AL157991; PIDN: CAB76210.1; GSPDB: GN00068
A:Experimental source: strain 972h(-); cosmid c24B10
C:Genetics:
A:Gene: SPDB:SPCC24B10.01
A:Map position: 3
C:Superfamily: replication licensing factor MCM4; MCM homology
F:1-301/Domain: MCM homology (fragment) <MCM>

Query Match 23.1%; Score 1064; DB 2; Length 407;
Best Local Similarity 55.9%; Pred. No. 2.1e-52;
Matches 221; Conservative 66; Mismatches 100; Indels 8; Gaps 4;
QY 503 INELEDIKKGLCOLPFGKAKKIPSGAS--FRGDINVLVDPGTSKSQLQVYVHKIAPR 560
Db 1 IYEMDDVKGLLLQLPFGGTNKSFGKASPRYRGDINILMCGDPSTSKSQLKYVHKIAPR 60
QY 561 GYITSGRGSSAVGLTAYVTKDPETRETIVLESALVLSDRGICCIDBFDKMSNARSLHE 620
Db 61 GYITSGRGSSAVGLTAYITRDQTKQLVLESALVLSDRGICCIDBFDKMSDATRSILHE 120
QY 621 VNEQQTYSVAKGGITASLNARTSVLACANPSGRYNARLSVIDNIOLPPTLLSRFDLYL 680
Db 121 VNEQQTYSVAKAGIITINARTSILASANPISGKYNPDLPTVKNIIDLPTLLSRFDLYL 180
QY 681 MLDKPEQNDRLARHLVALHYEN--YEVSKQDALDQLTAYITVARQHVHPTLSDEAA 738
Db 181 ILDRVDETLDRKLANHIVMYMBDTPHATDMEVFSVEFLTSYITARNINPVISEAA 240
QY 739 EDLNGYVEMROKGNPPGSKKVITATPROLESIMIRISALARMRSEVVEKVDAAEVR 798
Db 241 KELVNAVYGMKIGEDVRASEKRTITATPROLESIMIRLSAHAKHURNVVEGVDVLEAAR 300
QY 799 LLDVALQQSATHATGTIDMDLITTVGSASERIRANLLAALRELIADKISPGSSGLKT 858
Db 301 LIKTAIKDVTDPATGKISLDLIY--VNERETLVPEDMVKELANLISLNTVGGKT--MLV 356
QY 859 SOLLEDIRSQSSVDVSLQDIKNALGSLQGGFTV 893
Db 357 SOLLTRFQSSRLDASDFEACLGALERRGRIV 391

RESULT 10
H90226
minichromosome maintenance protein MCM [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: H90226
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, J.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: H90226
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-686 <KUR>
A:Cross-references: UNIPROT:Q9UXG1; UNIPARC:UPI0000066A8C; GB:AE006641; NID:gl3813947;
C:Genetics:
A:Gene: SS00774

Query Match 22.5%; Score 1039; DB 2; Length 686;
Best Local Similarity 34.4%; Pred. No. 1.2e-50;
Matches 244; Conservative 150; Mismatches 252; Indels 64; Gaps 15;
QY 195 FLHNYRSHADLNSKYIQIIBETVREEDTLMNDSIYDHDPLDYAKIVRYPLDIPIL 254
Db 17 FLTTFKGN--QNKYIERINELVAYRKSLIIEFSDVLSFNENLAYEINNTKIILPIL 74
QY 255 DTECOEVATSLPTPEKHIE---APFNLKASVHMRNLNPSDIDKLVSVKGMVIRCSII 311
Db 75 EGALYDHILQLDPTQYRDIEKVHVIRVIGIPRVIELRKIRSTDIGKLITIDGILVKTVPK 134
QY 312 PEIKGAFK----CLVCGHSP-----PLVTTVKGRVEPTCEKPECAARNAMSLIHR 361
Db 135 ERIYKATYKHHPDCMQBFEWPEDEMEPEV-----LEMPETIC--PKCKQKQOFLIPEK 186
QY 362 CTFANKQIVRLQETPDALPEGETPHTVSMCLYNTMVDVAVKPGDRIEVTGVFKAMAVRVGP 421
Db 187 TKLIDWQAVIQERPEEVPVSGQLPQLRIILLEDLVDLSARFGDRVKVTGILDIK--QDSP 244
QY 422 NORTLRALYKTYIDCVHVKSKDRGLQETDPMEMDKENDMYAGYHESDTSAAANEAKIOK 481
Db 245 VKRGSRAVFDIYMKVSSIEVSOK---VLDEVIISSEDE-----KK 281
QY 482 LKELSKLPGIDYRLSRSLAPSWELEDIKKGLLCLQFGKAKKIPSGASFRGDINVLVYG 541
Db 282 IKDLAKDPWIRDRIITSIAPSIYGHWEKELALALFPG--VPKVLIEDTRIRGDIHLIIG 340
QY 542 DPGTSKSQLQVYVHKIAPRGYITSGRGSSAVGLTAYVTKDPETRETIVLESALVLSDRGI 601
Db 341 DPGTAKSQLQVIRVAPRAVYTTGKSTAGLTAAVREKGTGEYLEAGALVADGGI 400
QY 602 CCIDBFDKMSNARSLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGRYNARLSV 661
Db 401 AVIDSIDKMRDSDRVAIHEAMEQQTVSIKAGIVAKLNARAIVAAAGNPKFGYISERP 460
QY 662 IDNIQIPPTLLSRFDLYLMLDKPQNDRLARHLVALHYENYEVSKQDALDQLTAY 721
Db 461 SDNINLPPTLLSRFDLYLILKDPQGEQ--DRELANYILDVHSGK---STKNIDIDILRKY 516
QY 722 IYARQHVHPTLSDEAADLNGYVEMROKGNPPGSKKVITATPROLESIMIRISALAR 781
Db 517 IYARQYVTPKITSKKNLITDFFVEMRKSSETDPS--LIITPRQLEALIRISAYAK 574
QY 782 MRFSEVVEKVDAAEVRLLDVALQQSATHATGTIDMDLITTVGSASERIRANLLAALR 841
Db 575 MALKARVTRDEAERAINMRFLFLESVGVDMESGKIDIDITMTGPKPSAREKMKIIEID 634
QY 842 ELIADKISPGSSGLKTSQLEDIRSQSSVDVSLQDIKNALGSLQGGFL 891
Db 635 SLAV-----SSECAKVKDILKEAQ---QVGIEKSNIEKLLTDMRKSGII 675

RESULT 11
D69103
DNA helicase (EC 3.6.1.-) MTH1770 [validated] - Methanobacterium thermoautotrophicum (B.
N)Alternate names: Cell division control protein CDC21/CDC54 homolog; replication licen
N)Contains: adenosinetriphosphatase (EC 3.6.1.3)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004

R;Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57541
A:Accession: S57553
A:Molecule type: DNA
A:Residues: 1-933 <BOW>
A:Cross-references: UNIPARC:UPI0000127256; EMBL:249919; NID:g887584; PIDN:CAA90164.1; PI
A:Experimental source: strain AB972
R;Conxon, A.; Maundrell, K.; Kearsey, S.E.
Nucleic Acids Res. 20, 5571-5577, 1992
A:Title: Fission yeast cdc21(+) belongs to a family of proteins involved in an early ste
A:Reference number: S26640; MUID:93087163; PMID:1454522
A:Accession: S26641
A:Molecule type: DNA
A:Residues: 571-646 <COW>
A:Cross-references: UNIPARC:UPI0000168852; EMBL:215032; NID:g3481; PIDN:CAA78750.1; PID:
C:Comment: The complex of six MCM proteins is one of several proteins that must be bound
phosphorylated and dissociate from the chromatin.
C:Genetics:
A:Gene: SGD.CDC54; HCD21; MIPS:YPR019W
A:Cross-references: SGD:S0006223; MIPS:YPR019W
A:Map position: 16R
C:Complex: The predominant form is a heterohexamer of MCM2 (PIR:S45757), MCM3 (PIR:A3637
component of replication licensing factor.
C:Function:
A:Description: MCM4 is a component of the replication licensing factor that permits DNA
C:Superfamily: replication licensing factor MCM4; MCM homology
C:Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
P:314-828/Domain: MCM homology <MCM>

Query Match 39.8%; Score 1833; DB 1; Length 933;
Best Local Similarity 40.9%; Pred. No. 5e-95;
Matches 393; Conservative 178; Mismatches 281; Indels 110; Gaps 14;

QY 2 ENNDALDIGAVSSPSQSEGVSTPLPQVTSFNDAAAPVAG----- 44
DB 11 EDNS-----SSPVVPDVP---PQLSSPALYSSSSQGDYGRNNSQLSQSGEN 61
QY 45 -RAVRQTP---TSAVRRRGRTDSARRRRSRSLGNSVSVYDAGTPGPGFPVATP 100
DB 62 IRAAIGSSPLNPFSSSQNSDVQSQGRQIRSSASAGRSRYSHLRSRALPTSS- 120
QY 101 VYATPVGTMGTPSFHRTGTPQYKQSELGSKPLHRRRSQSRFGRHSRSPESADGR 160
DB 121 -----SSLGRNQNRVHMRNDIHTSDLSSPRIVDFTDR 155
QY 161 -----PSEAEPPDTLGGVAYVVGTVNVAIPDVLRAIRFLHNYR----- 200
DB 156 SGVNTLDTSSSAPPSEASEP-----LRIWGTNVSIOECTTFRNPLMSFKYFKRKI 208
QY 201 -----SSAHLNSKYIITETVEREEDTLNIDMSDI--YDHPDLYAKIVRYPLDI 250
DB 209 LDREEFINTTDEELYIKQLENMRELGTSLNLDARNLLAYKQTEDLYHQLLYPQBV 268
QY 251 IPLLTECEQVATSL-----LPTPE-KHIEAPPNLKASVHMRNLNPSDIDKLVSV 300
DB 269 ISIMDQTIKDCMVSRVDNNLDYDDEIBETFKYKRYPNVVGCKGMRNLNPDIDKLINL 328
QY 301 KGMVTRCSSIIPKIGAPFKCLVCGHSPLVTVWVKRVEEPTKCEKPECAARNAMSLIHN 360
DB 329 KGLVLRSTVIEDMKVAFKCNVCDHT-MAVIDRGVIOEPARCEIDCNFNSSLIHN 387
QY 361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDVAVKPGDRIEVTGVFKAMAVRG 420
DB 388 RCSFADKQVILQETPDPVDPDQTPHSISLCVYDELVDSCRADRIEVTGTFERSIPIRAN 447
QY 421 PNQRTILALYKTYIIDCVHVKKSDRGRLQETDPEWMDKENDMTAGYHESDTSAAEAKIQ 480
DB 448 SRQRLKSLYKTYIVDVVHVKYSDKRLDVTSTIEQELMQNKVDHNEVESEVRQITDQDLA 507
QY 481 KLKELSKLPGIDRLSRSLAPSITWEEDIKGLLCLQFGGKAKIPSGASFRGIDNVLIV 540
DB 508 KIREVAAREDLISLLARSAPSIYELEDYKGIQLQLFGTNTKTFKGGRYGDNILILC 567

QY 541 GDPGTSKQLLYVHKIAPRGITYTSGRSSAVGLTAYVTKDPETRETIVLESGALVLSDRG 600
DB 568 GDPSTSKQLLYVHKITPRGVYTSKGSSAVGLTAYITRDYDKQLVLESGALVLSGG 627
QY 601 ICCIDEFDKMSDNARSMLEHVEOQTVSVAKGGIIASINARTSVLACANPSGRVNRALS 660
DB 628 VCCIDEFDKMSDSTRSVLHVEOQTVSIKAGIITTLNARSSILASANPIGRYNPNLP 687
QY 661 VIDNIQPLPTLLSPFDLIYMLDKPDEONDRRLARHLVALHYEN--YEVSKQDALDLQTL 718
DB 688 VTENIDLPPLLSRFDLYVLVDKDEKNDRELAKHLTNLYLEDKPEHISQDDVLPVBF 747
QY 719 TAYITARQHVHPTLSDEAAEDLNGVYEMKQGNFPGSSKKVITATPRQLESMLRISA 778
DB 748 TWIYSYAKHEHPIITEAKTELVRAYVGMKMGDDSRDEKRITATTRQLESMLRLAEA 807
QY 779 LARMRSEVVEKVDAAEARVLLDVALQOSATDHATGTIDMDLITTTGVSASERIRANLLA 838
DB 808 HAKMKLKNVVELEDVQEAVALRISAIKDYATDPKTGIDMNLVOTGKSVIQKQLQEDLSR 867
QY 839 ALRELIADKISPGSSSGLKTSQLEDIRSQSSVDVSLQIDIKNALGSLQSGELTYVHGDIV 898
DB 868 EIMNVLKDQ-----ASDSMSFNELIKQINEHSQDRVSSDIQEALSRLQOEDKVIVLGEGV 923
QY 899 KR 900
DB 924 RR 925

RESULT 4
S56766
replication licensing factor MCM4 - mouse
N:Alternate names: cell division cycle control protein CDC21/CDC54
C:Species: Mus musculus (house mouse)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: S56766
R:Kimura, H.; Takizawa, N.; Nozaki, N.; Sugimoto, K.
Nucleic Acids Res. 23, 2097-2104, 1995
A:Title: Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs and characteri
A:Reference number: S56766; MUID:95334361; PMID:7610039
A:Accession: S56766
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-862 <KM>
A:Cross-references: UNIPROT:P49717; UNIPARC:UPI00000041B9; EMBL:D26089; NID:g940405; PI
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1993
C:Comment: The complex of six MCM proteins is one of several proteins that must be bound
phosphorylated and dissociate from the chromatin.
C:Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C:Function:
A:Description: part of the replication licensing system that permits DNA replication to
C:Superfamily: replication licensing factor MCM4; MCM homology
C:Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
P:270-763/Domain: MCM homology <MCM>

Query Match 39.5%; Score 1821; DB 1; Length 862;
Best Local Similarity 43.7%; Pred. No. 2.1e-94;
Matches 401; Conservative 154; Mismatches 258; Indels 104; Gaps 18;

QY 30 VTSPSFDNAASPVAGRRVQR---TPTSAVRRRGRTDSARRRRSRSLG----- 77
DB 1 MSSP-----ASTP---SRRSRGRVTPQSLRSEERSSPNRRRRRSDSTGELLPMPTSP 54
QY 78 -----NSVYSSP-----YDAGTPGTPFVATPVVATPVGTPMGTP-8F 115
DB 55 GADLQSPAQNALFSSPPQMHSLAIPLDVSSPLT-----YGTSSRVEGTPRSG 105
QY 116 HRTGTPYQKQSELGSGQKPLHRRRSQSRFGRHSRSPESADGRPSSEABDDTLGGEY 175
DB 106 VRGTP-VRQRPDLGSAKGLQVDLOS-----DGAADIVPSEQSLGQK 148
QY 176 AVVWGTNNVPIPDVLRAIRFLHNY-----RSSAHLNSK-YIQIIBETVEREEDTLN 226

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:28:25 ; Search time 45 Seconds
(without alignments)
1926.473 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALIGAVSSPPSQS.....LGSLOGEGFLTVHGDIVKRV 901

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2332	50.6	720	2 C84540	probable CDC21 pro
2	1881	40.8	909	1 S26640	replication licens
3	1833	39.8	933	1 S56050	replication licens
4	1821	39.5	862	1 S56766	replication licens
5	1814.5	39.4	863	1 S65954	replication licens
6	1801.5	39.1	863	1 S64720	replication licens
7	1784.5	38.7	858	2 T47223	replication licens
8	1689.5	36.7	866	1 S59872	replication licens
9	1064	23.1	407	2 T50408	cdc21 protein [imp
10	1039	22.5	686	2 H90226	minichromosome mai
11	1020.5	22.1	666	1 D69103	DNA helicase [EC 3
12	1016.5	22.1	760	2 T39991	minichromosome mai
13	973	21.1	699	2 B72775	probable DNA repli
14	969	21.0	720	2 T47221	replication licens
15	966	21.0	719	1 JC4580	replication licens
16	963	20.9	775	1 A39531	replication licens
17	961.5	20.9	719	1 S70583	replication licens
18	961.5	20.9	881	2 T26498	hypothetical prote
19	954.5	20.7	720	1 A48723	replication licens
20	947.5	20.6	716	1 T01507	replication licens
21	947.5	20.6	735	1 PC4225	replication licens
22	940	20.4	936	2 E96508	hypothetical prote
23	938	20.4	586	1 E69314	replication licens
24	932	20.2	1017	1 S64219	replication licens
25	930	20.2	904	1 T10067	replication licens
26	928	20.1	734	1 I38080	replication licens
27	927.5	20.1	733	1 S56767	replication licens
28	927.5	20.1	759	2 T24130	hypothetical prote
29	924	20.0	868	1 S45757	replication licens

30	923	20.0	796	2 T47222	replication licens
31	920	20.0	810	2 H85865	protein ZK632.1 [i
32	914.5	19.8	892	2 T50339	mies protein [impo
33	910.5	19.8	868	2 T43423	probable replicati
34	909	19.7	845	1 S34027	replication licens
35	903.5	19.6	886	1 JC5085	replication licens
36	901.5	19.6	727	2 G84487	probable DNA repli
37	901.5	19.6	879	2 T41059	replication licens
38	897	19.5	807	1 I51685	replication licens
39	895.5	19.4	730	1 T03920	replication licens
40	882	19.1	830	1 B48723	replication licens
41	879.5	19.1	776	2 T52118	probable replicati
42	871.5	18.9	744	1 I51022	replication licens
43	858.5	18.6	971	1 A36376	replication licens
44	843.5	18.3	795	1 S22804	replication licens
45	842	18.3	812	2 T19446	hypothetical prote

ALIGNMENTS

RESULT 1

C84540

Probable CDC21 protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C:Accession: C84540

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84540

A>Status: preliminary

A:Residues: 1-720 <STO>

A:Cross-references: UNIPROT:Q9SIV8; UNIPARC:UPI00000A085A; GB:AE002093; NID:g4544386; P:

C:Genetics:

A:Gene: At2g16440

A:Map position: 2

Query Match 50.6%; Score 2332; DB 2; Length 720;

Best Local Similarity 61.0%; Pred. No. 3.3e-123; Indels 28; Gaps 6;

Matches 448; Conservative 124; Mismatches 135;

QY 154 EPSADGRPSESAPDDTLGGEYAYVWGTVNVPDVLRAIRRRFLHNY---RSSAHDL--NS 208

Db 5 EPLPSDDGDEEGDDTTP---TFVWGTVNISQDVKSAIEMFVKHFEARENSDDLFPREG 61

QY 209 KYIQIIEETVERBEDTLNIDMSDIYDHPDLVAKIVRYPLDIPLDITSCQEVATSLPPT 268

Db 62 KVMISIRKVIETEGEWIDVDAFDVDPDLNKNMVRYPLEVLAIAPDIVLMDIVSTINRL 121

QY 269 FKHTIETAPPNLKASVHRELNPSSIDKLVSVKGMVIRCSSLIPETIKGAFKFKCLVGHSP 328

Db 122 FEKHQVQRIIFNLTSTSMRNLNPSDIERKMSLKGMIIRSSSIPIPIREAVFRCVCGYFS 181

QY 329 PLVTVVKGVVEPTKCEKPECAARNAMSLIHNCRCTPANKQIVRLQETPDPAIPEGETPHTV 388

Db 182 DFIIVDRGKISPPPTCLKQECMTKNSMTLVHNCRFADKQIVRLQETPDPAIPEGETPHTV 241

QY 389 SMCLYNTMWDVAKPGDRIEVTGVFKAMAVRVGPNORTALYKTYIDCVHVKKSGDRGLQ 448

Db 242 SLLHDKLVNCGKPGDRIEVTGIRAMTVRVGPAHRTVKSPKTYIDCLHIKKAKLRMS 301

QY 449 TQDPMEMKXENDMYAGYHSDTSEANEAKIOKLKELSKLPGIYDLRSLSLAISELEL 508

Db 302 AEDPMDV-----DNSLRVDEDLRKQFQSLSPQDIYELRSLSLAPNIWELDD 348

QY 509 IKKGILLCOLFGGKAKKIPSGASFRGDINVLVGDPTGTSKSQLIQYVHKIAPRGYISGRG 568

Db 349 VKKGILLCOLFGGNALNLSAGNFRGDIINLLVGDPTGTSKSQLIQYIHKLSPRGIYISGRG 408


```
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 84
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-84

Query Match      2.8%; Score 128.5; DB 6; Length 2004;
Best Local Similarity 20.5%; Pred. No. 0.48;
Matches 200; Conservative 125; Mismatches 400; Indels 249; Gaps 47;

QY 17 PSQSEGVSTPLPQVTSFSDNAASPVAGRRVAVRQTPPTSAVRRRGRETDSARRRRSRSS- 75
DB 443 PVAGKRPDTVLP-VLNQVVAESAGRVSPKGMADAADFTRL-----ADRRRPEKAGV 496
QY 76 -LGNSVYSSPYD-----AGTGTGTPVATPVYATPVGTGTPGTFHRTGTPQYKORS 126
DB 497 PLGGGEYFEHTDRHIDALAGVPRPGKGMPEFADWAG-----PNSDGLVSDGRY 551
QY 127 ELGSGQKPLHRRRSQS--REPGH-----RSPS-----REPSADGRPSESAP- 167
DB 552 LKGREAEATLRAGGLSEAVPSEPRDYRPTQEARAPAKVMARPRDAADGKPAQRAQPARA 611
QY 168 -DDTLGGGYAYVGTWVNIIP--DVLRAIR-----RFLNHYRSSADLNSKYIQLIETVER 220
DB 612 KDTPVAGAAAANAATEKPSKDRNIEAGKSRFDGKGSAAGAAATEKPSSEKTKGA 671
QY 221 EEDTLNIDMSDIYDHPDLYAKIVRYPLDIIPLLQTEQEVATSL--PTPEKHIEARPF 278
DB 672 KPETFAKTASD-----NPE-----EARRKARVLOGGVYTVKEROAPQ 709
QY 279 NLKASVHMRLEINPSDIDKLIV-SVKGM-----VIRCSSIPE-----INGAFFKCLV 323
DB 710 GFKA---LREHAESIKKLAESIGLAERVDVAVSETAPDKAOMLLSQRVGEWF----- 761
QY 324 CHSPPLVTVVKGRVEEPTRECK---PECAARNAMSLIHNRTCTPAN-----KQIVR--- 371
DB 762 -----DGRTKITLVAENLTTPERAVMAAWHELHGRGFAADGFAKYREELERADG 810
QY 372 ---LQETPDALPEGE--TPHTVSMCLYNTWVDVAVKPGDRIEVT-----GVFKAMAVRGP 422
DB 811 NGLIRRIADAVOEGREGTGDAAASVRPAAVEAEVAELVAAQKTGGWAGIENRYGVKVGNG 870

; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 84
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-84

Query Match      2.8%; Score 128.5; DB 6; Length 2004;
Best Local Similarity 20.5%; Pred. No. 0.48;
Matches 200; Conservative 125; Mismatches 400; Indels 249; Gaps 47;

QY 17 PSQSEGVSTPLPQVTSFSDNAASPVAGRRVAVRQTPPTSAVRRRGRETDSARRRRSRSS- 75
DB 443 PVAGKRPDTVLP-VLNQVVAESAGRVSPKGMADAADFTRL-----ADRRRPEKAGV 496
QY 76 -LGNSVYSSPYD-----AGTGTGTPVATPVYATPVGTGTPGTFHRTGTPQYKORS 126
DB 497 PLGGGEYFEHTDRHIDALAGVPRPGKGMPEFADWAG-----PNSDGLVSDGRY 551
QY 127 ELGSGQKPLHRRRSQS--REPGH-----RSPS-----REPSADGRPSESAP- 167
DB 552 LKGREAEATLRAGGLSEAVPSEPRDYRPTQEARAPAKVMARPRDAADGKPAQRAQPARA 611
QY 168 -DDTLGGGYAYVGTWVNIIP--DVLRAIR-----RFLNHYRSSADLNSKYIQLIETVER 220
DB 612 KDTPVAGAAAANAATEKPSKDRNIEAGKSRFDGKGSAAGAAATEKPSSEKTKGA 671
QY 221 EEDTLNIDMSDIYDHPDLYAKIVRYPLDIIPLLQTEQEVATSL--PTPEKHIEARPF 278
DB 672 KPETFAKTASD-----NPE-----EARRKARVLOGGVYTVKEROAPQ 709
QY 279 NLKASVHMRLEINPSDIDKLIV-SVKGM-----VIRCSSIPE-----INGAFFKCLV 323
DB 710 GFKA---LREHAESIKKLAESIGLAERVDVAVSETAPDKAOMLLSQRVGEWF----- 761
QY 324 CHSPPLVTVVKGRVEEPTRECK---PECAARNAMSLIHNRTCTPAN-----KQIVR--- 371
DB 762 -----DGRTKITLVAENLTTPERAVMAAWHELHGRGFAADGFAKYREELERADG 810
QY 372 ---LQETPDALPEGE--TPHTVSMCLYNTWVDVAVKPGDRIEVT-----GVFKAMAVRGP 422
DB 811 NGLIRRIADAVOEGREGTGDAAASVRPAAVEAEVAELVAAQKTGGWAGIENRYGVKVGNG 870
QY 423 QR-----TLRALYKTYIDCVHVKKSDRGRLQTEDPMEM--DKENDW-----YAGY 465
DB 871 LKRGITAGVLARIGALLRRVLQRLAGKAG--GAMSADVPFAMLADLHGNVEGARDAPWGN 928
QY 466 HES-----DTSAAANEAKIQLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLQCLFG 520
DB 929 HRAVMFAAEADGAERSKSESLEKLRAETI--RISGREVPGGMLREYKRNAL----- 980
QY 521 KAKKIPSGAFSGDINVLVPGPGTSSQLLOLVYHKAIPRGYITSGRGSSAVGLTAYVTK 580
DB 981 -----EYGKSLRG-----PYVKNQDTGREI-SIGRS-----GITEILRH 1012
QY 581 DPETRETLESALVLSDRGICCIDFDMKSDNARSMLHEVMEQQTQVSVAKGGIITASLNA 640
DB 1013 DYKDAEHLQSIARIIQIENAVYIDTLPN-EDLAKNGDIOGYEYVYGLNVGG--ADYTV 1069
QY 641 RTSVLACANPSGRY-----NARLSVIDNTIQLPPTLSRFDLIYMLDKPQEND 690
DB 1070 RAAVAVSRN--GNRYVDHKLTKIEKGNLLSLDRVSTTCASESKPLSGI-----DD 1119
QY 691 RLARHLV-----ALHYENYEVSKQDALDLOTLTAYITVARQHVHTLSD----- 735
DB 1120 KRLLOILODKAGKGIADFTEAVRFSRAANIEAAGRITGCKSLRNALKDRWDASKG 1179
```

Db	705	AGELAVGREILKAGSIGVGVDYVUSCILGNAESNNWQELGRCLUI	753
Qy	691	RLRLAR-----HLVALHYENYVSKQADL--QTILYATTYARQHVPHTLSDEAAEDL	741
Db	754	RLLELTSABFYPHIVST--DVSQAPVPIRYKLLSLALFALQSI	798
Qy	742	INGVYEMROKGNFPCCSKVITATP--ROLESMTIRISE---ALARMPSVEVVEKYDAAE	795
Db	799	SHSWMGKLSRRYLLSSARWVITVPFLFSKLVTMLSSGSSHFAEMRRLMAIADEVIEAE	858
Qy	796	AVRL-----LD-----VAL-----QOSATDHATGTDMDLITTVGSASERRIRANLIA	838
Db	859	VTLQSGEDTLQOQDSSQALAPPYPRESSLEH-----TAHVETKGKGLKATRLUS	908
Qy	839	ALRELIADKISPGSSSGLKTSQLE	863
Db	909	ASSEDISDRLA-GVSVGLPSSATTE	932

RESULT 10

```

US-11-074-176-48
; Sequence 48, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences
; TITLE OF INVENTION: Stress-Related Proteins
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 420
; TYPE: Prt
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-48

```

Query Match	2.8%;	Score 129;	DB 7;	Length 420;
Best Local Similarity	22.2%;	Fred. No. 0.042;		
Matches	103;	Conservative 77;	Mismatches 164;	Indels 120; Gaps 23;

Qy	433	YIDCVHVKKSDRGRLQTEDPMEMDKNEMYAG---YHESDTSEAA-----NEAKIQKLKE	484
Db	5	PTDQEEIKCAFCGKTQ-----DQVKQMIAGNGVYICNECVDLAKKIIDDELRADSLKT	57
Qy	485	LSKLPFGIYDRLSRSLAPSIBWELEDIKGGLLQQLFG-GKAKTIPSGASF-----RGDI-	535
Db	58	ASELP-----KPEVEIKQLDQVYIGQDRAKKVLSVAVVHYHKRISQMDVD	102
Qy	536	-----NVLVGDGPGTSKQLQYVHKI--APRGIVTSGKSGSANGLT--AYVTKDPE	583
Db	103	SSTELQKSNIAMIGTSGKTVLAQTARILNVPPAI-----ADATTLTEAGYVGEDVE	156
Qy	584	TRETVLSEGA--LVLSDRGICCIDPEPKMSDNA-----RSMLEHVMEQQTVSVAKGGI	634
Db	157	NILLKLQGNADYDLERAQRGIYIDEIDKISKSENSVITRDVSGEGVQQSLKILEG--	214
Qy	635	IASLNARTSVLACPNPSGRYNARLSVDNIQLPPTLLSRFDLIYLM--LDKPDQONDR	691
Db	215	-----TTIASVPPQGRKHPQQEMIK-----MDTTNILEFVGAFDGIQIIVKS	257
Qy	692	RLARHLVALHYENVEVSKQDALDLQTLTAYITYA-----RQHVHPTTSLDEAA	738
Db	258	RLGKKTIGFGAEN-EVNKVQDADD---WTRHLHTADLVKFGMIPETGIRPIPIITTLDKLDN	313

```

QY 739 EDLINGVEMRGKNGFPFGSSKKVITATPQLESNMIIRISALARMRFSEVVVKYDAAEA VR 739
Db 314 KDLVRVLTE--PKNALVVKYKKLLSLDGVLEKFTDGALKAIADL-----AIORNMARGLR 367
QY 799 -----LLDVALQ-QSATHATGCTIDMDLITTVGSASERIRAN 835
Db 368 TTIENSIMDMYETPSEEDIESVEVTKDVITR--HAQPRITRKX 409

RESULT 11
US-10-467-657-7810
; Sequence 7810, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONGOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7810
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7810

```

Query Match	2.8%;	Score 129;	DB 6;	Length 498;
Best Local Similarity	20.4%;	Pred. No. 0.055;		
Matches	70;	Conservative 49;	Mismatches 122;	Indels 102; Gaps 11;
Qy	514	LCQLPGGKAKKIPSGASFRGDINVLVAGDPTGSKSOLLQVHKIAP-----	559	
Dd	193	LCDVKGGHSARLALEIAAAGSHLLMMGPPTGKSLMSQRDLGPLPLADELVEYWAALR	252	
Qy	560	-----RGYYTGRGSASVGLTAAYTVTKDPETRETIVLES GALVLSDRGICCIDCE	606	
Dd	253	SLLPNHQQQLDSNRPRSPHGHASAAWVG--GSDP-----	304	
Qy	607	FDKMDSNARSMLHEVMEOQTVSVAKGIIIASLNARTSVLACANP-----	650	
Dd	305	LPFDFRKVLVREPLENGEIHISRAARQAVYPAKFOLVAAAMPCCGYLGHPVKPCRCT	364	
Qy	651	--SGSRYNARLSVIDNIQLPPTLLSRPDLIYLMDKPDEQNDRRLABHVALHVNYEVFS	708	
Dd	365	PESVARYRSKIS-----GPLLDRID--LTIEVPS-----LSAAELMQOEAG	403	
Qy	709	KQDALDIQTLTAY--ITYARQ-VHVPITLS-----DEAAEDLINGVEMRQGNF	754	
Dd	404	ESSASVLERVIARNGKYARQGKVNAALSVELDSQACIQKEAQEAL-----	450	
Qy	755	PGSSKKYITATPQLESIMIRISEALARMRFPSEVVVKYDAABA V	797	
bh	451	-GSIALKVLSI-SARSFHRIMRVARTADLAGDEEVRSHRVNKAI	492	

RESULT 12

US-101-12
RSU-DI 467-657-94
; Sequence 84, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASINGANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GENOCOCCAL PROT

QY 882 L 882
Db 773 I 773

RESULT 5

US-11-087-227-16
; Sequence 16, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Malinowski, Timothy J.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-087-227-16

Query Match 18.9%; Score 870; DB 7; Length 543;
Best Local Similarity 37.0%; Pred. No. 8e-54;
Matches 200; Conservative 103; Mismatches 176; Indels 62; Gaps 13;
QY 317 AFFKCLVCG---HSPPLTVVKGVEPT-----RCEKPECAARNA---MSLIHNRCTPA 365
Db 4 ATYTCDQCAETYP-----IQSPTFMPLMCPQEQCTNRSGRLYLQTRGRPI 54
QY 366 NKQIVLQSTDAIPEGETPHPTVMCLNTYMTVDVAKPGDRIVTGTVFKAMAVRVGNPQRT 425
Db 55 KFQEMKQEHSDQVGNIPRISITVLVGENTRIAQPGDHVSVTGIFLPI-LRTGFRQV 113
QY 426 LRALYKTYIDCVHVKSDRGLQTEDPMEMKENDMYAGYHESDTSEANAELKJEL 485
Db 114 QGLLSETYLEAHRIVK-----MNKSEDESQ-----AGELTREELQI 151
QY 486 SKLPGIYDLRSLSAPSWELEDIKGLLCQLFGKAKKIPSGASFRGDIINVLLVGDQGT 545
Db 152 AE-EDFYEKLAASIAPEIYGHEDVKKALLLVGG-VQSPRGMKIRGNINICLMGDFGV 209
QY 546 SKSOLLQYVHKIAPRGITSGRSSAVGITA VYTKDPETRTVLESGALVLSDRGCCID 605
Db 210 AKSQLSTYIDRLAPRSQYTTGRSGSGVGLTAALVLRDSVSGELTLEGGALVLADQGVCCID 269
QY 606 EFDKMSDNARSMLEHVMEQQTVSVAKGIIIASINARTSVLACANPSGRYNARLSVIDNI 665
Db 270 EFDKMAEDRTAIHEVMEQQTISIAKAGLITLNAKCSILAAANPAYGRYNPRSLQNI 329
QY 666 QLPPTLSRFDLIYMLDKPDQNDRLARHVALHYENYVSKQ-DALDQTLTAYITY 724
Db 330 QLPALLSRFDLLIQLPRDRNDRLAQAHTYVHQHSRQPPSQFPLDMKLMRYIAM 389
QY 725 ARQVHPTLSDEAEDLINGYEMKQGNFPQSSKKVITATPROLESIMRISSEALARMF 784
Db 390 CREK-QPMVPESLADYITAAAYEMREAA---WASKDATYTSARTLTALIRLSALARL 445
QY 785 SEWKEVDAEAVRLDLVALQASATDHATGTDMDLITTVGSASERIRANLAALRELI 844
Db 446 VDVKEDVNEAIRLMEKSKSLGDD-----KGQTARTQPADVIFATVREL 493
QY 845 A 845
Db 494 S 494

RESULT 6

US-10-878-556A-80
; Sequence 80, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
; APPLICANT: Hoffmann La-Roche Inc.
; TITLE OF INVENTION: HCV regulated protein expression
; FILE REFERENCE: 21762
; CURRENT APPLICATION NUMBER: US/10/878,556A
; CURRENT FILING DATE: 2004-06-28
; NUMBER OF SEQ ID NOS: 199
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 80
; LENGTH: 808
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: sw_hum/mcm3_human
; DATABASE ENTRY DATE: 1992-05-01
US-10-878-556A-80

Query Match 18.3%; Score 845; DB 6; Length 808;
Best Local Similarity 27.4%; Pred. No. 8.1e-52;
Matches 232; Conservative 149; Mismatches 294; Indels 172; Gaps 18;
QY 180 GTNVNIPDV-LRAIRRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHPD 238
Db 3 GT-VVLDVLEAQRDYLDLDEED-QGIYQSKRELISDNQYRLVNVNDLRRKNEK 60
QY 239 LYAKIVRYPLDIIPLDTECOEVATSLPTFEKIEARPPNLKASVHMLNPSDIDK-- 296
Db 61 RANRLNNAPEELVAQFQALXDFVASIDATYAKQVEEYFVGLGSGFGSKHVSPTLTSCF 120
QY 297 ---LVSVKGMVIRCSSIPEIKGAFKCLVCGHSPPLTVVKGVEE----- 340
Db 121 LSCVVCVEGIVTKSLVRPKVRSVHYC-----PATKKTIERYSDLTTLVAFPPSSVY 174
QY 341 PTRCEKPECAARNAMSLIHNCRCTPANKOIVRLQETPDALPEGETPHTVSMCLINTMVDV 400
Db 175 PTKDEE-----NNPLETEYGLSVYKHOTITIQEMPEKAPAGQLPRSDVDLDDDLVDKA 229
QY 401 KPGDRIEVTGVFKAMAVRVGNPQRTLRALYKTYIDCVHVKSDRGLQTEDEPMEMDKEND 460
Db 230 KPGDRVQVGYRCLP---GKKGGVTSCTFTVLACNVK-----QMSK--- 270
QY 461 MYAGHESDTSEANAELKJELSKL--PGIYDLRSLSAPSWELEDIKGLLCQLF 518
Db 271 -----DAQPSFSAEDIAKIKFSTKTRSDIIPQLAKSLAPSIGHDYVKKAILCLLL 322
QY 519 GKGAKKIPSGASFRGDIINVLLVGDPGTSKSOLOLVVHKIAPRGIVTSGRSSAVGLTAVV 578
Db 323 GGVERDLENGSHIRGDIINILLIGPPSAKSQLRYVLTCTAPRAIPTTGSGSGVGLTAAV 382
QY 579 TKDPETRTVLESGALVLSDRGICCIDFDMKSDNARSMLEHVMEQQTVSVAKGIIIASL 638
Db 383 TTDQETGERRLAAGAWLADRGVVICIDFDMKSDMTAIHEVMEQGRVITAKAGIHARL 442
QY 639 NARTSVLACANPSGRYNARLSVIDNIQLPPTLSRFDLIYMLDKPDQNDRLRLARHLV 698
Db 443 NARCSVLAANPVGYRYDQYKTPMENIGLQDLSLSRFDLLFTMLDQMPEDQREISDVL 502
QY 699 ALH-----YENYEVSKODALDLQTLTAY----- 721
Db 503 RMHRYRPAPEGQGDAMPGLSAVDILATDDPNFSQBDQDTQIYEKHNDLLHGTCKKCKBM 562
QY 722 --ITYARQHVH-----PTLSDEAEDLINGVEMKQGNFPQSSKKVITATPROLESMI 773
Db 563 VSAAPMKYIIHAKIIRKPLVTOESATYIAEYSRLRSQDSMSSDTPARTSPVARTLETLLI 622
QY 774 RISEALARMFSEVVEKVDAAEAVRLDLVAL----- 804
Db 623 PLATAHAKARMSKTVLDQDAEENVELVQYAFKKVLEKEKKKKESEDESETEDEEKSQ 682

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:39:01 ; Search time 12 Seconds

(without alignments)
419.297 Million cell updates/sec

Title: US-10-768-511-8

Perfect score: 4609

Sequence: 1 MENNDALDIGAVSPSPSQS.....LGSQGBGLTVHGDIKRV 901

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 32527 seqs, 5584426 residues

Total number of hits satisfying chosen parameters: 32527

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:*

- 1: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960.5	20.8	719	7	US-11-087-227-14
2	943.5	20.5	904	7	US-11-087-227-12
3	930	20.2	904	6	US-10-507-275-5
4	916.5	19.9	821	7	US-11-087-227-90
5	870	18.9	543	7	US-11-087-227-16
6	845	18.3	808	6	US-10-878-556A-80
7	835	18.1	892	6	US-10-507-275-3
8	150	3.3	1493	7	US-11-004-057-4
9	145.5	3.2	1493	7	US-11-004-057-21
10	129	2.8	420	7	US-11-074-176-48
11	129	2.8	498	6	US-10-467-657-7810
12	128.5	2.8	2004	6	US-10-467-657-84
13	128.5	2.7	2004	6	US-10-467-657-6322
14	125.5	2.7	1299	6	US-10-821-234-1145
15	122	2.6	824	6	US-10-957-569-31
16	120.5	2.6	5024	6	US-10-793-626-2964
17	118	2.6	1249	7	US-11-039-398-22
18	118	2.6	1452	7	US-11-039-398-20
19	116	2.5	4384	6	US-10-821-234-1120
20	115	2.5	825	7	US-11-074-176-20
21	112	2.4	1095	6	US-10-793-626-3154
22	111	2.4	665	6	US-10-793-626-1232
23	111	2.4	1232	7	US-11-039-398-18
24	111	2.4	1235	7	US-11-039-398-16
25	110	2.4	1304	6	US-10-821-234-1648

Sequence 4910, Ap
Sequence 28, Appl
Sequence 200, App
Sequence 7004, Ap
Sequence 276, App
Sequence 2, Appl
Sequence 1129, Ap
Sequence 3988, Ap
Sequence 79, Appl
Sequence 4238, Ap
Sequence 6, Appl
Sequence 134, App
Sequence 956, App
Sequence 134, App
Sequence 4, Appl
Sequence 54, Appl
Sequence 49, Appl
Sequence 26, Appl
Sequence 47, Appl
Sequence 2, Appl
Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-11-087-227-14
; Sequence 14, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adrian J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-087-227-14

Query Match 20.8%; Score 960.5; DB 7; Length 719;
Best Local Similarity 33.2%; Pred. No. 5.6e-60;
Matches 236; Conservative 137; Mismatches 228; Indels 110; Gaps 21;

QY 192 IRRLHNYRSSAHDKNSKYIIEETV---BREEDTLNIDMSDIVDHPDL----- 239
Db 13 VKFKLQEFQD-DELGKKQFKYGNQVRLAHREQVALYVLDVDAEDDELVDISICENAR 71
QY 240 -YAKIVRYPL-DIIP-----LLDTECQ-----EVATSL 266
Db 72 RYAKLFADAVQELLQYKEREVVNVDVLYIEHRLMEQRGRDPGMRSPQNYPAELM 131
QY 267 PTFEKHIEARPNLKASVHMRLNPSDDIKLVSKGMVTCSSIIPEIKGAPFKCLVCG- 325
Db 132 RRFELYFQG-PSSNPRV-IREVRADSVGLVTVRGIVTVSEVKPMVVVATYTCQCGA 189
QY 326 --HSPPLTVTVKGRVEEPT-----RCEKPECAARNA---MSLIHNRCTFANKQIVRLQET 375
Db 190 ETYQP-----IQSPFTFMLIWCPSQECQTNRSQGRLYLQTRGSRIKFKEMKQEH 240
QY 376 PDPAIEGETPHTVYMCNTYNTVDVAKPGDRIEVTGVFKAMAVRVGNPQRTLRALYKTYID 435
Db 241 SDQVPVGNIPRSITVLVEGENTRIAPGDHVSVTGIFLPI-LRTGFRVQVQGLLSETYLE 299

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 6, 2005, 09:52:39 ; Search time 280 Seconds
(without alignments)
5719.939 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALDIGAVSPSPSQS.....LGSLLQEGFLTVGHGDIKRV 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-DB=issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511 @CGN 1 1.162 @runat 02122005 102335 18440 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/pp COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4609	100.0	4348	3	US-09-828-062-5
2	1139	24.7	723	3	US-09-828-062-2
3	993	21.5	2533	3	US-09-604-605-1
4	933	21.5	2533	3	US-10-101-080-1
5	982.5	21.3	3000	3	US-09-949-016-2621
6	946.5	20.5	3378	3	US-09-964-899-48
7	939	20.4	2533	3	US-09-949-016-5021
8	936	20.3	3445	3	US-09-976-594-323
9	918.5	19.9	2917	3	US-09-949-016-2468

10	856	18.6	3379	3	US-09-220-132-12	Sequence 12, Appl
11	817.5	17.7	1635	3	US-09-248-796A-4534	Sequence 4534, Ap
12	734.5	15.9	1488	3	US-09-248-796A-5904	Sequence 5904, Ap
13	670	14.5	996	3	US-09-248-796A-4570	Sequence 4570, Ap
c 14	656	14.2	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c 15	656	14.2	1664976	3	US-09-692-570-1	Sequence 1, Appli
16	474	10.3	1305	3	US-09-248-796A-4711	Sequence 4711, Ap
17	438.5	9.5	507	3	US-09-401-064-158	Sequence 158, App
18	427	9.3	12932	3	US-09-949-016-14363	Sequence 14363, A
19	391.5	8.5	519	3	US-09-248-796A-5916	Sequence 5916, Ap
20	381.5	8.3	1290	3	US-09-248-796A-4988	Sequence 4988, Ap
21	339	7.4	311	3	US-09-313-294A-83	Sequence 83, Appl
22	333	7.2	294	3	US-09-313-294A-4610	Sequence 4610, Ap
23	330.5	7.2	58407	3	US-08-916-421B-2	Sequence 2, Appli
24	330.5	7.2	58407	3	US-09-692-570-2	Sequence 2, Appli
25	312	6.8	297	3	US-09-313-294A-4978	Sequence 4978, Ap
26	289	6.3	2313	3	US-09-799-451-937	Sequence 937, App
27	266.5	5.8	613	3	US-09-533-559-658	Sequence 658, App
28	246.5	5.3	492	3	US-09-248-796A-5915	Sequence 5915, Ap
29	235	5.1	28366	3	US-09-949-016-16763	Sequence 16763, A
30	227	4.9	273	3	US-09-313-294A-1684	Sequence 1684, Ap
31	222.5	4.8	256	3	US-09-313-294A-2519	Sequence 2519, Ap
32	219	4.8	747	3	US-09-248-796A-4987	Sequence 4987, Ap
c 33	201.5	4.4	14636	3	US-09-173-914-6	Sequence 6, Appli
c 34	199	4.3	1059	3	US-09-248-796A-5905	Sequence 5905, Ap
c 35	195	4.2	13256	3	US-09-902-540-1006	Sequence 1006, Ap
c 36	189.5	4.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 37	188	4.1	565	3	US-09-328-111-454	Sequence 454, App
38	185.5	4.0	289	3	US-09-313-294A-3770	Sequence 3770, Ap
39	175.5	3.8	46899	2	US-08-471-119A-1	Sequence 1, Appli
40	169.5	3.7	28958	2	US-08-258-261B-6	Sequence 6, Appli
41	169.5	3.7	28958	2	US-08-456-837-6	Sequence 6, Appli
42	169.5	3.7	28958	2	US-08-457-342-6	Sequence 6, Appli
43	169.5	3.7	28958	2	US-08-457-646A-6	Sequence 6, Appli
44	169.5	3.7	28958	2	US-08-458-076A-6	Sequence 6, Appli
45	169.5	3.7	28958	2	US-08-764-233A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-828-062-5
; Sequence 5, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/09/828,062
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4348
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-062-5

Alignment Scores:
Pred. No.: 0
Score: 4609.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 3
Length: 4348
Matches: 901
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-768-511-8 (1-901) x US-09-828-062-5 (1-4348)

QY 1 MetGluAenAenAenAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
Db 1554 ATGGAAATAATGATGCACTTGACATTCGAGCCGCTGCTCCCATATCCTTCGCAATCT 1613
QY 21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaSer 40
Db 1614 GAAGAGTGTCTACGCCATTGCGCGCAAGTAACATCCAGGCCTCGCAATGCGACCTCA 1673
QY 41 ProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArgArgGly 60
Db 1674 CCGTGGCGCGGAGCGCGTACGCGCAGACCCCTACATCTGCAGTTTCGAAGGAGAGGG 1733
QY 61 ArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db 1734 AGAGAAACGGATTCGGCTCGTGTAGGAGAGTCGATCTCGGCTTTAGGCAATTCCTGTT 1793
QY 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db 1794 TATAGTTCCTTACGATGCGGGACTCTCGAATCTCTGGAATCTCCAGTGGCTACTCCG 1853
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db 1854 GTTTACGCTACCCAGTCGGTACACCTATGGTACCCCATCGTCCATCGTGGCAGCGCA 1913
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
Db 1914 CAGTACAACAGCGCAGTGAGCTTGGTTCCCGAGGGGAAGCCCTTACATCGGAGAGCTCGA 1973
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db 1974 TCTCAATCCAGAAACCCGGGCATCGATCTCTTCAGGGAACTTAGTGCTGATGGCGGT 2033
QY 161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
Db 2034 CCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAAATGCTTTATGTTTGGGGG 2093
QY 181 ThrAenValAenIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
Db 2094 ACGAATGTAACTTCAGATGTGCTTAGGGCGATTCGTCGAATTCCTCCACAATTTATCGT 2153
QY 201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluThrValGluArg 220
Db 2154 TCGAGTGTCTCATGATCTTAATTCGAATACATCCAGATCATAGAGAGACTGTGAGCGT 2213
QY 221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspIleAspProAspLeuTyr 240
Db 2214 GAGGAGGATACCTTAAATATCGACATGTACACATTTATGACCATGATCCTGATCTATAC 2273
QY 241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
Db 2274 GCAAAAATGTTCGATACCACTCGACATCATCCCCCTGTTGGACACATGAGTGTGAGAA 2333
QY 261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluLysAArgProPheAsnLeu 280
Db 2334 GTTGCTCTCTTACTACCAACGTTTCGAGAGCATATTTGAGGACGAGTTCGATCTC 2393
QY 281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db 2394 AAAGCATCGGTGCATCGGTGAATCAACCTTCAGATATAGACAAATTTGTTCTGTT 2453
QY 301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db 2454 AAAGAAATGGTTATCCGCTGAGTCTATCATACCTCGAATTTAGGGGGCCCTCTTCAAA 2513
QY 321 CysLeuValCysGlyHisSerProProLeuValThrValLysGlyArgValGluGlu 340
Db 2514 TGTITAGTGTGTGGTCACTCGCCTCCGCTAGTTTACAGTTGTTTAAGGGGGGTGAGGAG 2573
QY 341 ProThrArgCysGluLysProGluCysAlaAlaArgAenAlaMetSerLeuIleHisAsn 360
Db 2574 CCAACAAGGTGTGAAAGCCAGAAATGTGCAGCAGCAATGCTATGCTCTTTATTCACAA 2633

QY 361 ArgCysThrPheAlaAenLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db 2634 CGATGCACTTTTGCATAATAAGCAGATAGTGGTCTTCAAGAAACTCCAGATGCCATTCCT 2693
QY 381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db 2694 GAAGGAGAGACTCCACACAGTCAGCATGTGTTTATACACACATATGTTGATGCTGTG 2753
QY 401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db 2754 AAGCTCGAGATCGTATAGGTAAACAGGAGTTTCAAGGCGCATGGCAGTTCGAGTTGCT 2813
QY 421 ProAenGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db 2814 CCGAATCAACGAACATTACGAGCATTTGTAAGACCTTACATCGATTCGTCGACGCTCAG 2873
QY 441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db 2874 AAGTCTGACAGGGGTCACTGCAAACTGAAAGTCTTATGAGATGGATTAAGGAGAAATGAT 2933
QY 461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAenGluAlaLysIleGln 480
Db 2934 ATGTATGCTGGTATCATGAAAGTGATCTTCAAGAGCTGCTAATGAAGCAAAAGATTCAA 2993
QY 481 LysLeuLysGluLeuSerLysLeuProGlyTyrAspArgLeuSerArgSerLeuAla 500
Db 2994 AAATTAAGAGCTGTCCAGCTCCCGGCATTTATGATAGACTTTCAAGTCGCTGCT 3053
QY 501 ProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db 3054 CCAAGCATTTGGGAGCTTGAGATATTAAAGAGGCTTCTTTGCCAGCTCTTTGGTGGG 3113
QY 521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db 3114 AAGGCTAAGAAAATTCATCTCGAGCATCTTCCGAGGTGACATCAATGTTTACTTGT 3173
QY 541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db 3174 GGGAGCCCTGGTACAGTAATCTCAGCTGCTTATGTCAGTATGTGCACAAATAGCTCCTCGT 3233
QY 561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db 3234 GGAATCTACACTAGTGGCGAGGAGTTCCGGGTTGGGCTGACAGCTATGTAAACGAAG 3293
QY 581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db 3294 GATCCAGAACTCGAGAGACGCTATTCGAGAGCGAGCTTTGGTCTTCTAGTATCGTGGG 3353
QY 601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db 3354 ATATGCTGTATCGATGATTCGACAAATGCTGATTAATGCCCGAAGCATGCTTCATGAG 3413
QY 621 ValMetGluGlnGlnThrValSerValAlaLysGlyLysIleAlaSerLeuAsnAla 640
Db 3414 GTATGAGCAACAAACCGTATCTGTAGCCAAAGGGGGTATCATTTGCCCTCGCTGAACGCT 3473
QY 641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db 3474 CGGAGCTGTCTCTTCATGTGTGCAAAATCTAGTGGGTCCCGATCAATGCGCGCTTCT 3533
QY 661 ValIleAspAenIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db 3534 GTGATGATAACATCCAGCTTCTCCAACTCTACTCTCTAGATTTGATTTAATTTACTTTA 3593
QY 681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db 3594 ATGCTCGAACAACCGAGCAGCAAAACGATCGTCTCGCCAGGCATCTCGTGGCTTTA 3653
QY 701 HisTyrGluAenTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
Db 3654 CACTATGAAACTATGAATTTCAAAGCAGGACGCTTAGATCTACAAAACACTTACCCTG 3713

Qy 721 TyrIleThrTyraAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluasp 740
 Db 3714 TATATCACTATGCTCGTCAGCATGTACATCTACATTAAGTGAAGCTGCTGAAGAT 3773
 Qy 741 LeuIleAsnGlyTyrrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
 Db 3774 TTGATTAATGCTATGTTGAGATGCGCCAAAGGCAACTTCTCTGGGAAGCAGTAAAG 3833
 Qy 761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
 Db 3834 GTGATACAGCCACACCTCGGCNACTCGAAGATATGATTCGTATCAGTGAAGCCCTAGCT 3893
 Qy 781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
 Db 3894 CGAATGAGATTTCTGAAGTGTGAGAGAAAGTTGATGACAGCAGAAAGCTGTGCGCCTTTTA 3953
 Qy 801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
 Db 3954 GACGTGCGCTTTGACGAAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013
 Qy 821 IleThrThrGlyValSerAlaSerGluArgIleArgAlaAsnLeuLeuAlaAlaLeu 840
 Db 4014 ATCAGACTGGAGTGTGCGCCAGCAGCGATATTCGTGCGGCNACTTGTCTAGCTGCTCTG 4073
 Qy 841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
 Db 4074 CGAGAGCTTATAGCAGATATAAATTTTCACTGGCAGCTCTCTGCTTGAAGACCACTGAG 4133
 Qy 861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
 Db 4134 CTCTTGGAGATATCGGAGCCAAAGCAGTGTGGACGTTAGTTTTCAGGATATTAATAAT 4193
 Qy 881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
 Db 4194 GCTCTGGGTAGCTTCAAGAGAGAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253
 Qy 901 Val 901
 Db 4254 GTT 4256

RESULT 2
 US-09-828-062-2
 ; Sequence 2, Application US/09828062
 ; Patent No. 6710229
 ; GENERAL INFORMATION:
 ; APPLICANT: COSTA E SILVA, OSWALDO DA
 ; APPLICANT: BOHNERT, HANS J.
 ; APPLICANT: VAN THIELEN, NOCHA
 ; APPLICANT: CHEN, ROUYING
 ; APPLICANT: SARRIA-MILLAN, RODRIGO
 ; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
 ; FILE REFERENCE: 16313-0031
 ; CURRENT APPLICATION NUMBER: US/09/828,062
 ; CURRENT FILING DATE: 2001-08-20
 ; PRIOR APPLICATION NUMBER: 60/196,001
 ; PRIOR FILING DATE: 2000-04-07
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 723
 ; TYPE: DNA
 ; ORGANISM: Physcomitrella patens

Alignment Scores:
 Pred. No.: 5,74e-87 Length: 723
 Score: 1139.00 Matches: 226
 Percent Similarity: 98.26% Conservative: 0
 Best Local Similarity: 98.26% Mismatches: 3
 Query Match: 24.71% Indels: 1
 DB: 3 Gaps: 0

US-10-768-511-8 (1-901) x US-09-828-062-2 (1-723)
 Qy 403 GlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsn 422
 Db 10 GGAGATCGTATTTGAGGTAAACAGAGATTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAAT 69
 Qy 423 GlnArgThrLeuArgAlaLeuTyrrLysThrTyrrLysAspCysValHisValLysLysSer 442
 Db 70 CAACGAAACATTTACGAGCATTTGTATAAGACCTTACATCGATTGCGTGACGTCGAAGAGTCT 129
 Qy 443 AspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrr 462
 Db 130 GACAGGGGTCCACTGCAAACTGAAGATCCTATGGAGATGGATTAAGAGAAATGATATGTAT 189
 Qy 463 AlaGlyTyrrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeu 482
 Db 190 GCTGGGTATCATGAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAAGATTCAAAAACCTT 249
 Qy 483 LysGluLeuSerLysLeuProGlyIleTyrrAspArgLeuSerArgSerLeuAlaProSer 502
 Db 250 AAAGAGCTGTCCAAGCTCCGACATTTTATGATAGACTTTCAGAGTCTGCTGGCTCCAAGC 309
 Qy 503 IleTyrrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAla 522
 Db 310 ATTTGGGAGCTTGAAGATATTAAGAGGCTCTTTCTTCCAGCTCTTTGGTGGAGAGGCT 369
 Qy 523 LysLysIleProSerGlyValaSerPheArgGlyAspIleAsnValLeuLeuValGlyAsp 542
 Db 370 AAGAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTACTTCTTGGGGAC 429
 Qy 543 ProGlyThrSerLysSerGlnLeuGlnTyrrValHisLysIleAlaProArgGlyIle 562
 Db 430 CCTGTACAGTAATCTCAGCTGCTCAGTATGTGTGCAAGATAGCTCTCTGCTGGAAATC 489
 Qy 563 TyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrrVal-ThrLysAspPr 582
 Db 490 TACACTAGTGGCGAGGAAGTTTCGGCGGTGGGCTGACAGCGCTATGTAAACGAGGATCC 549
 Qy 582 oGluThrArgGluThrValLeuGluSerGlyValaLeuValLeuSerAspArgGlyIleCy 602
 Db 550 AGAAACTCGAGAGACGGTATTTGGAGCGGAGCTTTGGTCTTAGTGATCGTGGGATATG 609
 Qy 602 sCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMe 622
 Db 610 CTGTATCGATGAGTTTCGACAAATGCTGTATAATGCCGAGCATGCTTCATGAGGTAAT 669
 Qy 622 tGluGlnThrValSerValaLys 631
 Db 670 GGAGCAACAAACGGTATCTGGACCCAAG 697

RESULT 3
 US-09-604-605-1
 ; Sequence 1, Application US/09604605
 ; Patent No. 6421613
 ; GENERAL INFORMATION:
 ; APPLICANT: Ramgopal Nadimpalli
 ; APPLICANT: Carl R. Simmons
 ; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
 ; FILE REFERENCE: 1138
 ; CURRENT APPLICATION NUMBER: US/09/604,605
 ; CURRENT FILING DATE: 2000-06-27
 ; PRIOR APPLICATION NUMBER: 60/143,222
 ; PRIOR FILING DATE: 1999-07-09
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2533
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (110) ... (2269)
 US-09-604-605-1


```
QY 740 spLeuileAsnGlyTyrValGluMetArgGln-----LysGlyAsnPheProGlyS 757
   |||
   |||
   |||
Db 1839 ACATTGCAACTGCATATTCACGATCCGCGAGAGGCGGAGTCAAAATGCACCAACT 1898
   |||
   |||
   |||
QY 757 erSerLysValysValleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerG 777
   |||
   |||
   |||
Db 1899 CC-----TACCAACTATCAGAACACCTTTTGAGCATACTCCGATATCTCTA 1943
   |||
   |||
   |||
QY 777 luAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlav 797
   |||
   |||
   |||
Db 1944 TTGCCTTGGCAAGACTGAGTTCTCAGAACTGTGGCTCAGAGCGATGTCAGAGGAC 2003
   |||
   |||
   |||
QY 797 alArgLeuLeuAspValala-----LeuGlnGlnSerAlaT 809
   |||
   |||
   |||
Db 2004 TGCAGCTGATGCAAAATGTCAGTACTCATTTACTCAGATGACCGCGCTCTGGCC 2063
   |||
   |||
   |||
QY 809 hrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerG 829
   |||
   |||
   |||
Db 2064 TTGATGCAATATCTGACATATATTCATCTCAGAGACGAAGCTGCTAGGACGAGCA 2123
   |||
   |||
   |||
QY 829 luArgIleArgAlaLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleS 849
   |||
   |||
   |||
Db 2124 TGGATGTGAGATACGCTCAGCTCTTAACCTGATC----- 2158
   |||
   |||
   |||
QY 849 erProGlySerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnS 869
   |||
   |||
   |||
Db 2159 -----TCCAGAAAGGATACAGCGAGCTCAATTGAAGCAATGCTTGGAGGAATACG 2210
   |||
   |||
   |||
QY 869 erSerValaspVal 873
   |||
   |||
   |||
Db 2211 CTTCCCTGAATGTG 2224
```

RESULT 4

```
US-10-101-080-1
; Sequence 1, Application US/10101080
; Patent No. 6879916
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/10/101,080
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/604,605
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110)...(2269)
US-10-101-080-1
```

```
Alignment Scores:
Pred. No.: 9.8e-74 Length: 2533
Score: 993.00 Matches: 281
Percent Similarity: 49.88% Conservative: 131
Best Local Similarity: 34.02% Mismatches: 292
Query Match: 21.54% Indels: 122
DB: 3 Gaps: 25
```

US-10-768-511-8 (1-901) x US-10-101-080-1 (1-2533)

```
QY 84 ProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
   |||
   |||
   |||
Db 9 CCTTCTCCCGGCATCCCGGAACACTTGGCCGCCCCCTCCCGCTCCCGCTGTTCTCT 69
   |||
   |||
   |||
QY 104 Thr-ProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLy 123
   |||
   |||
   |||
```

```
Db 69 GCCTCCCAACCCCGAGCGG-----CCGTCCCATCACCGCCCTGGGATGAAGAACC 119
QY 123 s-----GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSe 141
   |||
   |||
   |||
Db 120 CTGACTTCGCGCGAGACAAAGCGCTCGCAAGGACTTCTGTCCAATTTACCGG----- 174
   |||
   |||
   |||
QY 141 rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr 161
   |||
   |||
   |||
Db 175 -----CCCTCAGCGGAGGCCAAGTACCTGGAACATCC 206
   |||
   |||
   |||
QY 161 oSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGlyTh 181
   |||
   |||
   |||
Db 207 TGCAGATGTGCGAACAGGAAGATACGCGGCTCCAGATCGAGCT-----GGAGC 257
   |||
   |||
   |||
QY 181 rAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSe 201
   |||
   |||
   |||
Db 258 ACCTATTCCACTATTAAGGATGTCGACGAGGAGTTCTTCAGCGGTGTCCGAGAACACGC 317
   |||
   |||
   |||
QY 201 rSerAlaHis---AspLeuAsnSerLysTyr-IleGlnIleIleGluGlu----- 216
   |||
   |||
   |||
Db 318 GCGCTACATCGGCATCTTCGCGGAGGCCATGACGAGATCATGCGGAGGCCACAGAGG 377
   |||
   |||
   |||
QY 217 -----ThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspH 235
   |||
   |||
   |||
Db 378 CQTACACCGTCCAGAGGACCCAGACATCCTG-----ATGACGAGCGTGTGGATG 428
   |||
   |||
   |||
QY 235 sAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuA 255
   |||
   |||
   |||
Db 429 AGGGGCGCGAGCGCGCGGAC-----GGCAGCGACCCGCTA----- 466
   |||
   |||
   |||
QY 255 spThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleGluA 275
   |||
   |||
   |||
Db 467 -----CAGAAATATGCCACGAGATCAAAACGGTCTTCTGAGGTTTACATCAAG 515
   |||
   |||
   |||
QY 275 laArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleA 295
   |||
   |||
   |||
Db 516 CA-----TTCTCAAAGGTGACTCCACTCACCTTCGGCAAGTGAAGGATCAACATTTG 569
   |||
   |||
   |||
QY 295 spLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleL 315
   |||
   |||
   |||
Db 570 GTCAGTTGTGAAATATCTGGATTTGTTACTCGCTCGCTCGGATGGAAGCCCTTGATGC 629
   |||
   |||
   |||
QY 315 ysGlyAlaPheLysCysLeuValCysGlyHisSerProProGluValThrValValL 335
   |||
   |||
   |||
Db 630 AGGTTGCTGTTTATATCATGTGAAGAATGTGGCTTCGAG-----ATATACCAAGAA 683
   |||
   |||
   |||
QY 335 ysGlyArgValGluGluPro---ThrArgCysGluLysProGluCys-----AlaA 351
   |||
   |||
   |||
Db 684 CTGCTAGAGTGTATTATGCCCTTCATTGAATGTCATCTCAACGATCAAGCTGAACAAAG 743
   |||
   |||
   |||
QY 351 laArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValA 371
   |||
   |||
   |||
Db 744 CGAAGGGAATCTAATCTTCACTCGGTCGCTCAAAAGTTTGTGAATTTTCAGGAGTAA 803
   |||
   |||
   |||
QY 371 rgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetC 391
   |||
   |||
   |||
Db 804 AGCTTCAAGAGTTAGCAGAGCACGTACCAAGAGGCCACATTCCTCGTCTCTGACTGTC 863
   |||
   |||
   |||
QY 391 ysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyV 411
   |||
   |||
   |||
Db 864 ATCTAAGAGGAGAGCTGACTAGAAAGTTGCACTCGAGATGTGGTTGATGTGCGGTA 923
   |||
   |||
   |||
QY 411 alPheLysAlaMet-----AlaValArgValGlyProAsnGlnArgT 425
   |||
   |||
   |||
Db 924 TTTTCTCCTATGCCATCTACTAGGGTTTAGGCGGATGCGCGAGA----- 970
   |||
   |||
   |||
QY 425 hrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal-----HisValLysLys 442
   |||
   |||
   |||
Db 971 -----TTAGTTGCTGATACTTACTTGGAAACAATGTGTGTACCCTCACTTCAAGAAA 1022
   |||
   |||
   |||
QY 442 erAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetT 462
   |||
   |||
   |||
Db 1023 AA----- 1024
```

```
QY 462 yrAlaGlyThrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysL 482
Db 1025 -----TATGAGGAGTATGCTTAAAGGTGATGAACAAG-----CAAA 1064
QY 482 euLysGluSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProS 502
Db 1065 TTGACCGATTGGCTGAGGACGGTGATATCTACAGTAAGCTGGCAAGGTCTCTGGCACCTG 1124
QY 502 erIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysA 522
Db 1125 AAATATTGGCCATGAAGATGTCAAAAAGGCACTGCTGTACTACTCTGTGTGTGTCACCCC 1184
QY 522 laLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyA 542
Db 1185 ATCGGAAGCTCGCAGATGGCATGAAGATCAGAGGAGACCTGCATATGTCATGATGGGAG 1244
QY 542 spProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyI 562
Db 1245 ATCTGGTGTGCAAAAGAGTCACTTCTGAAGCATATTATCANTGTTGCACCNAGAGGAG 1304
QY 562 leTyrThrSerGlyArgGlySerSerAlaValGlyLeuThraAaTyrValThrLysAspP 582
Db 1305 TGTACACCACTGGACGTGGAGCAGTGGTGTGGACTTACCGTCCAGTCCAGAAAGATC 1364
QY 582 roGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleC 602
Db 1365 CAGTAAACAATGAGTTTCTCTCGAAGTGGAGCACCTGGTATCGGCAGATATGGGCATTT 1424
QY 602 ysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValM 622
Db 1425 GTGCTATAGATGAGTTTACACAGATGGAGAGTCAAGAGCAGACGACGATTCATGAGGTAA 1484
QY 622 etGluGlnGlnThrValSerValAlaLysGlyGlyIleAlaSerLeuAsnAlaArgT 642
Db 1485 TGGAGCAGCAACACGTAGCATTCGCAAGGCTGGCATCACCACTCTCTTAATGCGAGAA 1544
QY 642 hrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValI 662
Db 1545 CTGCAATTTCTGGCTGCTGCTCAATCAGCATGGGGAAGGTATGATATGAGAGAACCCAG 1604
QY 662 leAspAsnIleGlnLeuProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetL 682
Db 1605 CAGAAAATATAAATCTACCTCAGCTCTTCTGCTCGTTTCAGCTCTTGGTTAACTCC 1664
QY 682 euAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisT 702
Db 1665 TGGATCGTCGACACATGGAACCTGATCTTGAATGGCAAGACACGTTGTTTCATGTGCAT- 1723
QY 702 yrGluAsnTyrGluValSerLysGln-----AspAlaLeuAspLeuGlnThrLeuThra 720
Db 1724 --CAAAATCTTGAATCACCAGCGCTGGGGTTCCACACCACTTGAGCCATCTGTACTCAG 1781
QY 720 laTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluA 740
Db 1782 CATACATATCTGCTGTAGAGA---GTCATTCCTTCTGTTCTCTAGAGAGCTCGAGGAAT 1838
QY 740 spLeuIleAsnGlyTyrValGluMetArgGln-----LysGlyAsnPheProGlyS 757
Db 1839 ACATTGCAACTGCATATTCAGCATCCGCCAAGAGGAGCGAA3TCAAATGCACCAACCT 1898
QY 757 erSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerG 777
Db 1899 CC-----TACACAACATATCAGAACAACCTTTGAGCATATCTCCGTATCTCTA 1943
QY 777 luAlaLeuAlaArgMetArgPheSerGluValGluLysValAspAlaAlaGluAlav 797
Db 1944 TTGCCTTTGGCAAGACTAGGTTTCTCAGAAACTGTGGCTCAGAGCGATGTCCAGCAGAC 2003
QY 797 alArgLeuLeuAspValala-----LeuGlnGlnSerAlaT 809
Db 2004 TGCAGCTGATGCAAAATGTCCAAGTACTCATTTATCTCAGATGATGACCGCAGCGGTCTGCC 2063
```

```
QY 809 hrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerG 829
Db 2064 TTGATGCAATATCTGCATATATTCATCTCTGAGAGAGCAAGCTGTAGGAGCAGCA 2123
QY 829 luArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleS 849
Db 2124 TGGATGTGAGATACGCTCACGCTCTTAACTTGATC----- 2158
QY 849 erProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnS 869
Db 2159 -----TCCAGAAAGGATACACGAGGCTCAATTGAGGAATGCTTGGAGGAATACG 2210
QY 869 erSerValAspVal 873
Db 2211 CTTCCCTGAATGTG 2224

RESULT 5
US-09-949-016-2621
; Sequence 2621, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2621
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2621
```

```
Alignment Scores:
Pred. No.: 9,97e-73 Length: 3000
Score: 982.50 Matches: 299
Percent Similarity: 45.41% Conservative: 156
Best Local Similarity: 29.84% Mismatches: 304
Query Match: 21.32% Indels: 244
DB: 3 Gaps: 36
```

US-10-768-511-8 (1-901) x US-09-949-016-2621 (1-3000)

```
QY 67 ArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAsp 86
Db 7 CGGAACAAAGAACGCGTGTGCGTGGCCCTTTAAGAGCGA-TTCTCTCCGCCCGCGCC 65
QY 87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValAlaThrProVal 106
Db 66 AGCTCGGACCGCGGGAACCGCGCGCTGCACTACCCGCCCGGAGATTCCCTCCGAGC 125
QY 107 GlyThrProMetGlyThrProSerPheHisArgGlyThr----- 119
Db 126 CCGGACCGCTCCCGTCACTCATCTTAGGCCCGCAGCGTATGGCTTGGCGGTAGCG 185
QY 120 -----ProGlnTyrLysGlnArgSerGlu 127
Db 186 GAGGTGAAGAAGCGCGCTTGTCCGATTGGCCCGCAGCAGTGGCGC---CGGTCACT 242
QY 128 LeuGly-----SerGlnGlyLysProLeuHisArgArgArgSer 141
Db 243 GGGGGCGGACGTTTTCGCCCAATTTCGGTTGGCCGCCACAGTCCACCGCGCGAGATTC 302
QY 142 GlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPro 161
```



```

Qy 745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysValleThrAla 764
Db 2388 TACGTGGAGATGAGCGAGAGGCT-----TGGGCTAGTAAGATGCGCACCTATACT 2438

Qy 765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
Db 2439 TCTGCCCGACCTGCTGCTATCTGCGCTTCCACTGCTTGGCAGCTGAGAAATG 2498

Qy 785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
Db 2499 GTGATGTGTGTGGAGAAAGAGATGTGAATGAAGCCATCAGGCTAATGGAGATGTCAAAG 2558

Qy 805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGly 824
Db 2559 GACTCTCTTCTAGGAGAC-----AAGGGG 2582

Qy 825 ValSerAlaSerGluArgIleArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle 844
Db 2583 CAGACAGCTAGGACTCAGAGACCAGCAGATGTGATATTGGCCACGTCGCTGAACTG--- 2639

Qy 845 AlaAspLysIleSerProGlySerSer-SerGlyLeuLysThrSerGlnLeuLeuGluAs 864
Db 2640 -----GTCTCAGGGGGCGAAGTGTCCGGT-----TC 2666

Qy 864 PileArgSerGlnSerSerValAspVal-----SerLeuGlnAs 877
Db 2667 TCTGAGGAGAGCAGCAGCTGTGTATCTCTGTGGCTTTCACACCGGCCAGATTCCAGCGGCT 2726

Qy 877 PileLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIle 897
Db 2727 CTGGATGATATAGGAGCTCAATCTTGGCAGGTCAATGTCTCCGGACACGGATCACT 2786

Qy 897 eVal 898
Db 2787 TTGTG 2790

RESULT 6
US-09-964-899-48
; Sequence 48, Application US/09964899
; Patent No. 6900367
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al
; TITLE OF INVENTION: Identification of Genes Involved in
; FILE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 3378
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-964-899-48

Alignment Scores:
Pred. No.: 1,34e-69 Length: 3378
Score: 946.50 Matches: 262
Percent Similarity: 46.45% Conservative: 157
Best Local Similarity: 29.05% Mismatches: 344
Query Match: 20.54% Indels: 135
DB: 3 Gaps: 23

US-10-768-511-8 (1-901) x US-09-964-899-48 (1-3378)

Qy 29 GlnValThrProSerPheAspAsnAlaAlaSerProValAlaGlyArgAlaVal 48
Db 4 CAGGAATCATCGAATCTCTTACCATGGCATCCAGCCCG---GCCAGCGTGGCGAGGC 60

```

```

Qy 49 ArgGlnThrProThrSerAlaValArgArgGlyArgGluThrAspSerAlaArgArg 68
Db 61 AATGATCTCTCACTCCAGCCCTGCGCGAAGCTCCCGCGGTACTGATGCGCTC----- 114

Qy 69 ArgArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGly 88
Db 115 ACCTCGAGCCCTGCGCTGACCTT-----CCACCATTTGAGGATGAG 156

Qy 89 ThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThr 108
Db 157 TCCGAGGGGCTCTTAGGCACA-----GAGGGGCGCTTGAGGAAGAAGAGATGAGAG 210

Qy 109 ProMet-----GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArg 125
Db 211 GAGCTCATTTGAGATGGCATGGAAGGAGACTACCGCGCCATCCAGAGCTGGACGCCCTAT 270

Qy 126 SerGluLeuGly-----SerGlnGly 132
Db 271 GAGCGCGAGGAGCTGGCTCTGGATGATGAGACAGCTAGAGGAGCTGACGGCCAGTCAGAGG 330

Qy 133 LysProLeuHisArgArgArgSerGlnSerArgGluProGly----- 147
Db 331 GAGGACAGAGCGGCGCATGCGCAGCGTGACCGGAGGCTGCGCGGGCGCTGGCGCGC 390

Qy 148 -----HisArgSerProSerArgGlu 154
Db 391 ATGCGCGCTGGGCTCTGTATGACAGCGATGAGGAGGAGGAGGAGCGCCCTGCCCGCAAG 450

Qy 155 ProSerAlaAspGlyArgProSerGluSerAlaGluProAsp----- 168
Db 451 CGCGCGAGGTGGAGCGGCGCACGAGGACCGGAGGAGGAGGAGGAGGAGATGATTGAGAGC 510

Qy 169 -----AspThrLeuGlyGlyGluTyrAlaTyrValTrp-----Gly 180
Db 511 ATCGAGAACCTGGAGGATCTCAAAGGCCACTCTGTGCGCGAGTGGCTGAGCATGCGCGG 570

Qy 181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--- 199
Db 571 CCGCGGCTG-----GAGATCCACCCCGCTTCAAGAACTTCTCTCGCAGCTCACGTC 621

Qy 200 ArgSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGlu 219
Db 622 GACAGCCACGCCCAACGTC-----TTCAAGGAGCGCCATCAGGACATGTGCAAA 672

Qy 220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db 673 GAGAACCGTGAGAGCCTGGTGAACCTATCAGGACTTTGGCAGCGCCAGGAGGACGCTGCTG 732

Qy 240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db 733 GCCTACTTCTCTGAGGACCGCGGAGCTGCTGCAGATCTTTGATGAGGCTGCCCTG 792

Qy 260 GluValAlaThrSerLeuLeuProThrPheGluLys-----HisIleGluAlaArg 276
Db 793 GAGTGTGTACTGGCCATGTACCCCAAGTACACCGCATCACCACCATCATCATGCTCGC 852

Qy 277 ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLys 296
Db 853 ATCTCCACCTGCTCTGTTGGAGAGCTGCTGCTGAGGAGCTGCATCGATGAACCCAG 912

Qy 297 LeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGly 316
Db 913 CTGATCCGCAACAGTGGGGTGTGCCAGCTGTCATGCGCTCTGCGCCAGCTCAGCATG 972

Qy 317 AlaPhePheLysCysLeuValCysGlyHisSerProProLeuValIleThrValValLysGly 336
Db 973 GTCAGTACAACCTGCAACAGTGAATTCCTCTGGGTCTCTTCTGCGAGTCCCGAGAC 1032

Qy 337 ArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSer 356
Db 1033 CAGGAGGTGAACACAGGCTCTCTGT-----CTGAGTCCAGTCCGCGCGGCTTTGAG 1086

Qy 357 LeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrPro 376

```

```
Db 1087 GTCAACATGGAGGAGACCATCTATCAGAACTACAGAGGTATCCGAATCCAGGAGAGTCCA 1146
Qy 377 AspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMet 396
Db 1147 GCGAAGTGGCGGCTGGCGGCTGCCCGCTCCAAGGAGCCGCAATTCCTCCGACATCTG 1206
Qy 397 ValAspAlaValIleProGlyAspArgIleGluValThrGlyValPheLysAlaMetAla 416
Db 1207 GTGGACAGCTGCAAGCCAGGAGACGAGATAGAGCTGACTGGCATCTATCAACAAC----- 1260
Qy 417 ValArgValGlyProAsnGlnArgThrLeuArgAla-----LeuTyrLysThrTyrIle 434
Db 1261 ---AACTATGATGGCTCCCTCAACACTGCCAATGGCTTCCTCTGTCTTTGCCACTGTGCATC 1317
Qy 435 AspCysValHisValLysLysSerAspArg-----GlyArgLeuGlnThrGlu 450
Db 1318 CTAGCCAAACACCGTGGCCCAAGAGACACAAGGTGTGTAGGGGAAGCTGACCGATGAA 1377
Qy 451 AspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThr 470
Db 1378 GAT----- 1380
Qy 471 SerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGly 490
Db 1381 -----GTGAAGATGATCCTCCTCAAGGATCAGCAG 1416
Qy 491 IleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrPheGluLeuGluAspIleLys 510
Db 1417 ATCGGAGAGAAAGATCTTTGCCAGCATGTCTCTCCATCTATGATGATGATGATGATGATGAT 1476
Qy 511 LysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer 530
Db 1477 AGAGGCTGGCTGTGGCTGTGGAGGGGAGCCCAAAACCCAGGTGGCAGCAGCAG 1536
Qy 531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550
Db 1537 GTACGTGGTGATCAACGTGCTTGTGCGGAGACCTTGGCAGCAGCAAGTCCGAGTTT 1596
Qy 551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570
Db 1597 CTCAAAGTATATTGAGAAAGTGTCCAGCCGAGGCATCTTCCACCTGGCCAGGGGGCGTGC 1656
Qy 571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
Db 1657 GCTGTGGGCTCAGCGCGTATGTCAGCGGACCCCTGTGACAGGGAGTGGACCTTGGAG 1716
Qy 591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610
Db 1717 GCTGGGCGCTGTTCTGCTGACCGAGGAGTGTCTCATGATGATGATGATGATGATGATGATG 1776
Qy 611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnThrValSerValAla 630
Db 1777 AATGACCGAGACAGAACCAAGCATCCATGAGGCGCATGAGGAGCAGCAGCATCTCCATCTCG 1836
Qy 631 LysGlyGlyIleLeuAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnPro 650
Db 1837 AAGGCTGGCATGCTACCTCCCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCG 1896
Qy 651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThr 670
Db 1897 ATAGGAGGGCGGTACAGCCCTCGCTGACTTCTCTGAGAGAGTGGACCTCAGAGGCC 1956
Qy 671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
Db 1957 ATCATCTCAGCTTTGACATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2016
Qy 691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
Db 2017 GAGATGCTGGCGGCTTCTGTGGTGGGCGAGCCAGTCCAGACACCACCCCGACCAAGGAG 2076
Qy 710 ----- 710
```

```
Db 2077 GAGGAGGGGCTGGCCCAATGGCAGCGCTCTGAGCCCGCATGCCCAACACGATATGGCGTG 2136
Qy 711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrAlaArgGlnHisValHis 730
Db 2137 GAGCCCTTGGCCAGGAGGTCTTGAAGAAGTACATCATCTAGCCCAAGGAGGGGTCCAC 2196
Qy 731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
Db 2197 CCGAAGCTCAACAGCATGGACCAAGGTGGCCAAAGATGTACAGTGCACCTGAGGAAA 2256
Qy 751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
Db 2257 GAATCTATGGCGCAGCGCAGC-----ATCCCATTTACGTGGCGCAGCATCGAG 2304
Qy 771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
Db 2305 TCCATGATCCGATGGCGGAGGCCCAAGCGGCATCCATCTCGGGGACTATGTGATCGAA 2364
Qy 791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db 2365 GACGACGTCAACATGCCATCCGCGTG-----ATGCTGGAGAGCTTCATAGAC 2412
Qy 811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArg 830
Db 2413 ACACAGAAGTTCAGCGTCTATGCGCAGCATGCGCAAGACT---TTTGCCCGCTACCTTTCA 2469
Qy 831 IleArgArgAlaAsn-----LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLys 847
Db 2470 TTCCGGCGTGACACAAATGAGCTGTGTCTTCTCATCTAGACGAGTGTAGTGGCAGAGCAG 2529
Qy 848 IleSer 849
Db 2530 GTGACA 2535
```

RESULT 7

```
US-09-949-016-5021
; Sequence 5021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5021
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5021
```

```
Alignment Scores:
Pred. No.: 3,61e-69 Length: 2533
Score: 939.00 Matches: 232
Percent Similarity: 51.71% Conservative: 156
Best Local Similarity: 30.93% Mismatches: 290
Query Match: 20.37% Indels: 72
DB: 3 Gaps: 18
```

US-10-768-511-8 (1-901) x US-09-949-016-5021 (1-2533)

```
Qy 169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db 119 GACAGCTTCGGGGCGAC---GCCCGAGGCGGAGGCGGCGCCGCCAAATCGCAGCTG 175
```

189 LeuArgAlaIleArgArgPheLeuHisAsnTyrArg-----SerSerAlaHisAspLeu 206
176 CAGAGCGCTTCAAGAGTTCTCGGCAGATACCGAGTGGCCACCGACCGCGGCTTC 235
207 AsnSerLysTyrIleGlnIleIleGluGluThrValGluArg3LugluAspThrLeuAsn 226
236 ACCTTCAAATACAGGAGTAACTCAAGCGCATTAACAACCTG3GGGAGTACTGGATTGAG 295
227 IleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyr 246
296 GTGGAGATGGAGATCTCGCCAGCTTTGATGAGGACCTGGCC3ACTACTTGTACAAAGAG 355
247 ProLeuAspIleProLeuLeuAspThrGluCysGlnGluValAla-----Thr 263
356 CAGCGGAGCACCTCGACGTCTGGAGGAAGCTGCCAAGAGGTAGCTGATGAGGTGACC 415
264 SerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSer 283
416 CGGCGCCCGCCTTCTGGGAGGAGGTGCTCCAGGACATCCAGTCTCAAGTCCGAC 475
284 Val-----HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
476 GCCAGCCCTTCCAGCATCTGAGCTGAAAGTGGGACATGATCTCACACCTGGTGAAGATC 535
301 LysGlyMetValIleArgCysSerIleProGluIleLysGlyAlaPhePheLys 320
536 CTTGGCATCATCATCGCGCCTCTCGGTTCGTGCAAGGCCACCGCATCTCTATCCAG 595
321 CysLeuValCysGlyHisSerProLeuValThrValIleLysGly-----ArgVal 338
596 TGCCGCGAGTCCGCAACACACCTTCCACCATTTGCCATGGCCCTGGCGCTAT 655
339 GluGluProThrArgCysGlu-----LysProGluCysAlaAlaArgAsn 353
656 GCCCTGCCAGGAAGTGAACACAGATCAGGTGGCGGCCCAATGCCCCATTG---GAC 712
354 AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln 373
713 CGGTACTTCATCATCCCGACAAATGCAATGCGTGGACTTCCAGACCTGAAAGTGCAG 772
374 GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr 393
773 GAGTGCCTGATGAGTCCCCACCGGAGATGCCACACATCGACGTCTACTGCGAC 832
394 AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
833 AGGTACCTGTGTGACAAAGTCTGCTGGGAACAGGTATCCATCATGGGCATCTACTCC 892
414 AlaMetAlaVal-----AtqValGlyProAsnGlnArg 424
893 ATCAAGAAGTTTGGCCTGACTACACAGCGGGGCCGTGACAGGGTGGCGTGGCGATCCGA 952
425 ThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArg 444
953 AGC-----TCCTACATCCGTGCTCGGCGATCCAG-----982
445 GlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGly 464
983 -----GTGGACACAGATGGCTCTG3CCGCGAGCTTTCTGGG 1018
465 TyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGlu 484
1019 -----GCCGTGAGCCCCCAGGAGGAGGAGGTTCGCTCGC 1054
485 LeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrp 504
1055 CTGGCTGCCCTCCAAATGCTATGAGTCACTCCAGAGCATCGCCCCCTCCATCTTT 1114
505 GluLeuGluAspIleLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLys 524
1115 GGGGCGCACAGATGAAGAAGCCATTCCTCCCTGCTCTTTGGGGGTCCCGAAAGAGG 1174
525 IleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuValGlyAspProGly 544

1175 CTCCTGATGACTTACTCGCCGAGGAGACATCAACCTGCTGATGCTAGGGGACCCTGGG 1234
545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
1235 ACAGCAAGTCCCAGCTTCTGAAGTTTGTGAGAAGTGTCTCCCATTTGGGTATACAGC 1294
565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
1295 TCTGGGAAGACAGCAGCGCAGCTGGAGCTGACAGCTCGGTGATGAGGAGCCCTTCGTCC 1354
585 ArgGluThrValLeuGluSerGlyValAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
1355 CGGAATTCATCATGGAGGCGGAGCCATGCTCTGGCCGATGGTGGGGTCTCTGTATT 1414
605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
1415 GACGAGTTTGACAAGATGCGAGAAGATGACCGTGTGGCAATCCACGAAGCCATGGAGCAG 1474
625 GlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVal 644
1475 CAGACCATCTCTATCGCAAGCTGGGATCACCAACCCCTGAATCCCGCTGCTCGCTC 1534
645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
1535 CTGCTGCTGCTCAACTCAGTGTTCGGCCGCTGGGATGAGACGAGGGGAG---GACAAAC 1591
665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
1592 ATTGACTTTCATGCCACCACCATCTTGTGGCGCTTCGACATGATCTTCATCGTCAAGATGAG 1651
685 ProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
1652 CACAATGAGGAGAGGATGTGATGTGCCAAGCATGTCTCATCTCTGCACGTGAGCGCA 1711
705 TyrGluValSerLys-----GlnAspAlaLeuAspLeuGlnThrLeuAlaTyrIle 722
1712 CTGACACAGACACAGGCTGTGGAGGCGGAGATGACCTGGCCAGCTCAAGAAGTTTATT 1771
723 ThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIle 742
1772 GCCTACTGCCGAGTGAAGTGTGGCCCGCGCTGTCCAGCAGAGGCTGCAGAGAACTGAAG 1831
743 AsnGlyTyrValGluMetArg-----GlnLysGlyAsnPheProGlySerSerLys 759
1832 AACCGCTACATCATCATGCGGAGCGGGCCCGTTCAGCACGAGAGGACAGTGCACCGCCG 1891
760 LysValIleThrAlaThrProArgGlnLeuGlnSerMetIleAtqIleSerGluAlaLeu 779
1892 TCCAGCATCCCATCATCTGTCGGCAGCTGGAGGCCATTTGGCATTCGCGGAACCCCTC 1951
780 AlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeu 799
1952 AGCAAGATGAAGTGCAGCCCTTCGCCACAGAGGCAGATGTGGAGGAGGCCCTCGCGCTC 2011
800 LeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAsp 819
2012 TTCCAAGTGTCCAGTTGGATGCTGCC-----TTGTCCGGTACCTCTG-----2053
820 LeuIleThrThrGlyValSerAlaSerGluArgIleArgAlaAsnLeuLeuAlaAla 839
2054 -----TCAGGGGTGGAGGGTTCACCGCCGAGGAGGACCGAGGATGCTGAGCCGC 2104
840 LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSer 859
2105 ATCGAGAAGCAGCTCAAGCGCGCTTTGCCATTTGGCTCCAG-----GTGCTCAGCAGC 2158
860 GlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLys 879
2159 AGCATCATCAAGGACTTCCAAAGCAAAATATCCCGGAGCAGCCATCCCAAGGTGTGTG 2218
880 AsnAlaLeu-----GlySerLeuGln 886


```
Db 1522 GACATCAAGAGAGGCGCTGCTGCTGCGCCCTGTTCCGAGGGAGGCCCAAAAAACCAGGTGGC 1581
Qy 528 GlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLys 547
Db 1582 AAGCAAGGTACGTGGTGATATCAACGTCTTGTGCGGAGACCCCTGGCAGCAGGAAG 1641
Qy 548 SerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
Db 1642 TCGCAGTTCTCAAGTATATTGAGAAAGTGTCCAGCGGCACCTGTGTCCAGCAGGGAGTGG 1701
Qy 568 GlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr 587
Db 1702 GGGGGCTGGCTGTGGCTCACGGCTATGTCTCAGCGGCACCTGTGTCCAGCAGGGAGTGG 1761
Qy 588 ValLeuGluSerGlyAlaIleuValLeuSerAspArgGlyIleCysIleAspGluPhe 607
Db 1762 ACCTTGGAGGCTGGGGCCCTGGTCTGCTGACCGAGGAGTGTGTCTCATTTGATGAATTT 1821
Qy 608 AspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVal 627
Db 1822 GACAAGATGAATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1881
Qy 628 SerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCys 647
Db 1882 TCCATCTCGAAGGCTGGCATGCTACCTCCCTGCGAGGCTGCTGCGACGCTCATTTGCTGCC 1941
Qy 648 AlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeu 667
Db 1942 GCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAACGTTGACCTC 2001
Qy 668 ProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGlu 687
Db 2002 ACAGAGCCCATCATCTACGCTTTGACATCTGCTGTGTGGTGGGACACCGTGGACCCA 2061
Qy 688 GlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVal 707
Db 2062 GTCCAGGACGAGATGCTGGCCGCTTCTGTGGTGGCGAGCCACGTCAGACACCCACCCAGC 2121
Qy 708 SerLysGln----- 710
Db 2122 AACAGGAGGAGGAGGGGCTGGCCAAATGGCAGCGCTGTGAGCCGCGCATGCCCAACACG 2181
Qy 711 -----AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGln 727
Db 2182 TATGCGGTGGAGCCCTGCCCCAGGAGTCTCTGAAGAAGTACATCATCTACGCCCAAGGAG 2241
Qy 728 HisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGlu 747
Db 2242 AGGGTCCACCCGAAGCTCAACGATGGACGACGACGACGACGACGACGACGACGACGACG 2301
Qy 748 MetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArg 767
Db 2302 CTGAGGAAGAATCTATGCGGACAGGACG-----ATCCCATTTACGCTGGCGG 2349
Qy 768 GlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluVal 787
Db 2350 CACATCGAGTCCATGATCCGATGCGGAGGCGCCAGCGCGCATCCATCTGCGGGACATAT 2409
Qy 788 ValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSer 807
Db 2410 GTGATCGAAGACGACGTCACATGCGCATCCCGGTG-----ATGCTGGAGAGC 2457
Qy 808 AlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrGlyValSerAla 827
Db 2458 TTCTATAGACACACAGAAGTTACGCGTCATGCGCAGCATGCGCAAGACT---TTTGCCGCG 2514
Qy 828 SerGluArgIleArgArgAlaAsn-----LeuLeuAlaAlaLeuArgGluLeuIle 844
Db 2515 TACCTTTTCATTCGCGCGTGACAAATAGAGCTGTGCTCTTCACTAAGCAGGTAGTG 2574
Qy 845 AlaAspLysIleSer 849
Db 2575 GCAGAGCAGGTGACA 2589
```

RESULT 9

```
US-09-949-016-2468
; Sequence 2468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2468
; LENGTH: 2917
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2468
```

Alignment Scores:

Pred. No.:	2,46e-67	Length:	2917
Score:	918.50	Matches:	267
Percent Similarity:	45.43%	Conservative:	121
Best Local Similarity:	31.26%	Mismatches:	301
Query Match:	19.93%	Indels:	165
DB:	3	Gaps:	24

US-10-768-511-8 (1-901) x US-09-949-016-2468 (1-2917)

```
Qy 124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGlnSer 143
Db 33 CGCAGGCGAATCTCGGCACCTAAGCAATATGGACCTCGCGCGCAGCGAGCGCGGG 92
Qy 144 -ArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerG 163
Db 93 CGCGCAGCCAGCACCTCGGAGGTCCGCGACGA-----GGTCCCGAGAGAGTGCC 140
Qy 163 uSerAlaGluPro-AspAspThrLeuGlyGlyGlyTyrAlaTyrValTrpGlyThrAsnV 183
Db 141 AGAAACTCTTCTCTGGAC----- 157
Qy 183 alAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerA 203
Db 158 -----TTCTTGAGGAGGTTTTCAGACGAGC- 181
Qy 203 laHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluA 223
Db 182 ----GATGGAGAAAATAAATACTTGCAATATAGCAGAGGAACTGATTCTGCTCGAGAGAA 236
Qy 223 spThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysI 243
Db 237 ACACATTGGTTGTGAGTTTGTGGACCTGGAACAATTTAACCCAGCAACTTTCCACCACCA 296
Qy 243 leValArgTyrProLeuAspIleIleProLeuAspThrGluCysGlnGluValAlaIat 263
Db 297 TTCAAGAGGAGTTCTATAGAGTTTACCTTTACCTG-----TGTCCG----- 337
Qy 263 hrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPhe----- 278
Db 338 ----GCCTTGAAACAATTCGTAAGACCGTAAGAGATCCCTCTTGCAAGGATTTT 392
Qy 279 -----AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspI 294
Db 393 ATGTTGCAATCCAAAGACCTGCTACGACACACAGATTCGAGAGCTCACCTCATCCAGAA 452
Qy 294 leAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluI 314
```


Db 453 TTGGTTTGTCTCACTCGCATCGTGGCAGGTGGTGGGACTCACCCAGTTCACCCAGAGC 512
Qy 314 lEylsGlyAlaPhePheLysCysLeuValCysGlyHisSerProLeuValThrValV 334
Db 513 TTGTGAGCGGAATCTTTCTGCTTGGACTGT-----CAGACAGTGA 554
Qy 334 aLlysGly-----ArgValGluGluProThrArgCysGluLysProGluC 349
Db 555 TCAGGGATGTAGAACAGCAGAGTTCAAAATACACACACCAACATCTGCCGAATCCAGTTT 614
Qy 349 yAlaAlaArgAenAlaMetSerLeuLeuHisAenArgCysThrPheAlaAenLysGlnI 369
Db 615 GTGCCAACAGGAGAGATCTTACTGTGATACAAATAAATCAAGATTGTGTATTTTCAA 674
Qy 369 lEValArgLeuGlnGluThrProAenAlaLeuProGluGluGluThrHisThrValS 389
Db 675 AGGTTCGTATTCAGAGAGCCCAAGCTGAGCTTCTCGAGGGAGTATCCCCCGCAGTTTGA 734
Qy 389 eRMetCysLeuTyraenThrMetValAspAlaValLysProGlyAspArgGleGluValT 409
Db 735 AGTAATTTTAGGGCTGAAGCTGTGGATCAGCTCAAGCTGGTGACAGTGTGACTTTA 794
Qy 409 hRgLyVal-----PheLysAlaMetAlaValArgValG 420
Db 795 CAGGACACACTGATTGTGTGCTGCTCTCAAGCTTAGCACACACGAGCAGCAGTGCAG 854
Qy 420 lYProAenGlnArgThr-----LeuA 427
Db 855 AAATAATTTCCCGTGTCACTGTGTGATGGATATGAGACAGAGGATTCGAGGACTCC 914
Qy 427 iGAlaLeu-----TyLys---ThrTyLysCysValHisValL 440
Db 915 GGGCCCTTGGTGTAGGAGACCTTCTTATAGGCTGCTTCTTCTGCTGCTGT- 967
Qy 440 yElysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnA 460
Db 968 -----GTTGGCCCAACCAACC 983
Qy 460 sPMetTyralaGlyTyRHISGluSerAspThrSerGluAlaAlaAenGluAlaLysIleG 480
Db 984 CAAGGTTTGGGGGAAGAGCTCAGAGATGAGAACACAGACAGCTGAGACATTTAAGAAC 1043
Qy 480 lN-----LysLeuLysGluLeuSerLysLeuProGlyIleTyA 493
Db 1044 AATGACTGTGAAGAAATGGGAGAAAGTGTGTGATGATGAGTCAAGTAAATAATCTATACC 1103
Qy 493 sPArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyL 513
Db 1104 ACAATCTTTGTACCAGCCTTTCCTTACTATACATGGCAATGATGAAGTAAACGGGGTG 1163
Qy 513 euLeuCysGlnLeuPheGlyLysAlaLysLysIleProSerGlyAlaSerPheArgG 533
Db 1164 TCCTGTGTGTCTTGTGGTGGGTTCCAAAGACACACAGGAGAGGACCTCTCTTCGAG 1223
Qy 533 lYAspIleAenValLeuValGlyAspProGlyThrSerLysSerGlnLeuGluLnt 553
Db 1224 GGGACATAATGTTTCATTTGTTGGTGACCAAGTACAGTAAAGAGCAATTTCTCAAGC 1283
Qy 553 yRValHisLysIleAlaProArgGlyIleTyThrSerGlyArgGlySerSerAlaValG 573
Db 1284 ACCTGGAGGAGTTTCAGCCCGCAGAGCTGTCTACACCAATGGTAAAGCGTCCAGTGTCTG 1343
Qy 573 lYLeuThrAlaTyValThrLysAenProGluThrArgGluThrValLeuGluSerGlyA 593
Db 1344 GCTTAACAGCAGCTGTTGTGAGAGATGAAGAAATCTCATGATTGTTCATTGAGGCTGGAG 1403
Qy 593 lAlaValLeuSerAspArgGlyLysCysIleAspGluPheAspLysMetSerAspA 613
Db 1404 CTTTGTATGTGGCTGATAATGGTGTGTGTATGATGAATTTGATGAATGAGATGAGCTGC 1463
Qy 613 snAlaArgSerMetLeuHisGluValMetGluGlnThrValSerValAlaLysGlyG 633

Db 1464 GGGATCAAGTTGCTATTTCATGAAGCTATGGAAACAGCAGACCATATCCATCTCAATAAGCG 1523
Qy 633 lYlLeAlaSerLeuAenAlaArgThrSerValLeuAlaCysAlaAenProSerGlys 653
Db 1524 GAGTGAAGCTTACTCTGAACCGCCGAGCTCCATTTTGGCAGCAGCAAAACCAATCAGTG 1583
Qy 653 eRArgTyAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThrLeuLeuS 673
Db 1584 GACACTATGACAGATCAAAATCATTTGAACACAGAATATAAATTTGTGAGTCCCATCATGT 1643
Qy 673 eRArgPheAspLeuIleTyLeuMetLeuAspLysProAspGluGlnAenAspArgG 693
Db 1644 CCGGATTCGATCTCTTCTTATCTTGTGGATGAATGAATGAGTTTACAGATTATGCCA 1703
Qy 693 euAlaArgHisLeuValAlaLeuHisTyRgluAenTyRgluValSerLysGlnAspAlaL 713
Db 1704 TTGCCAGCGCATAGATTTCAT---TCAAGAATTTGAGGAATCAATTTGATCGTGTCT 1760
Qy 713 euAspLeuGlnThrLeuThrAlaTyIleThrTyAlaArgGlnHisValHisProThrL 733
Db 1761 ATTCCCTCGATGATATCAGAAAGATATCTTCTCTTTCGAAGACAG---TTTAAACCCAAGA 1817
Qy 733 euSerAspGluAlaAlaGluAspLeuIleAenGlyTyRValGluMetArgGlnLysGlyA 753
Db 1818 TTTCAAAGAGTCAGAGGACTTCATTGTGGAGCAATATAAATCATCTCCGCCAGAGATG 1877
Qy 753 snPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetI 773
Db 1878 GT---TCTGGAGTGACCAAGTCTTCATGGAGGATTCAGTGGCAGACCTTTGAGAGCATCA 1934
Qy 773 lEArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspA 793
Db 1935 TTTGCTCTCTGAAGCTATGGCTCGGATGCTGCTGTGATGAGGTCCAAACCTAAACATG 1994
Qy 793 lAlaGluAlaValArgLeuLeu-----AspValA 803
Db 1995 TGAAGAGACTTTCCGGTTACTGTAATAAATCAATCATCGTGTGGAAACACCTGATGTCA 2054
Qy 803 lAlaGlnGlnSer-----AlaThrAspHisAlaThrGlyThrIle---- 816
Db 2055 ATCTAGATCAGAGGAAGAGATCCAGATGGAGGTAGATGAGGGTGTGTGGCANTCAATG 2114
Qy 817 -----AspMetAspLeuIleThrThrGlyValSerAla----- 827
Db 2115 GTCATGTGTGACGCCCTGTCTCTGTGAACGGATCAATGGCTACAATGAAGACATAAATC 2174
Qy 828 -----SerGluArgIleArgArgAlaA 835
Db 2175 AAGAGTGTCTCCCAAGCCTCTTAAAGCTGGGCTTCTCTGAGTACTCCCGAATCTCTA 2234
Qy 835 snLeuLeuAlaAla---LeuArgGluLeuIleAlaAspLysIleSerProGlySerSers 854
Db 2235 ACCTTATTTGTGTTCCACTCAGAAAGGTGGAAGAAAGAG-----GACGAGT 2282
Qy 854 eRgLyLeuLysThrSerGlnLeu-----LeuGluAspIleArgSerGlnSers 870
Db 2283 CAGCATTAAAGAGGAGCGAGCTTTGTTAACTGGTACTTTGAAGAAATCGAATCAGAGTAG 2342
Qy 870 eRValAspValSerLeuGlnAspIleLysAsnAlaLeu 882
Db 2343 ACTCTGAAGAGAACTTATAAATAAAAAAAGAATCATA 2380

RESULT 10

US-09-220-132-12

; Sequence 12, Application US/09220132

; Patent No. 6506607

; GENERAL INFORMATION:

; APPLICANT: Shvlian, Andrew W.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT

; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANC

; FILE REFERENCE: 07334-074001

; CURRENT APPLICATION NUMBER: US/09/220,132

; CURRENT FILING DATE: 1998-12-23

; PRIOR APPLICATION NUMBER: US 60/079,303
 ; PRIOR FILING DATE: 1998-03-25
 ; PRIOR APPLICATION NUMBER: US 60/068,821
 ; PRIOR FILING DATE: 1997-12-24
 ; NUMBER OF SEQ ID NOS: 191
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 3379
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-220-132-12

Alignment Scores:

Pred. No.:	6,038-62	Length:	3379
Score:	856.00	Matches:	244
Percent Similarity:	45.26%	Conservative:	143
Best Local Similarity:	28.54%	Mismatches:	322
Query Match:	18.57%	Indels:	147
DB:	3	Gaps:	19

US-10-768-511-8 (1-901) x US-09-220-132-12 (1-3379)

```

QY 31 ThrSerProSerPheAspAlaSerProValAlaGlyArgAlaValArgGln 50
DB 13 TCATCGGAATCTTCACCATGGCATCCAGCCG---GCCAGCTGTCGGGAGCAATGAT 69
QY 51 ThrProThrSerAlaValargArgGlyArgGluThrAspSerAlaArgArgArg 70
DB 70 CCTCTCACCTCCAGCGCTGGCGAAGCTCCCGCGCTACTGATGCCCTC-----ACCTCC 123
QY 71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
DB 124 AGCCCTGGCGGACCTT-----CCACCATTTGAGGATGAGTCCGAG 165
QY 91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
DB 166 GGGCTCTTAGGCACA-----GAGGGGCCCT--GGAGGAAGA 203
QY 111 GlyThrProSerPheHis-----ArgGlyThrProGlnTyrLysGlnArg 125
DB 204 GGATGGAGAGGAGCTCATTTGGAGATGGATGGAAGGAGACTACCGCGCATCCAGAGCT 263
QY 126 SerGluLeu-----128
DB 264 GAGCGCTATGAGCGCGAGGAGCTGGCTCTGGATGATGAGGACGTAGAGAGCTGACGGC 323
QY 129 -----GlySerGlnGlyLysProLeuHis-----136
DB 324 CAGTCGAAGGAGGAGCAGCAGCGGCCATCGCGCACGGTACCGGGAGCTGGCGGGGCT 383
QY 136 -----136
DB 384 GGGCGCATGCGCGTGGGCTCTGTATGACAGCGATGAGGAGCAGGAGCGCCCTGCC 443
QY 137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
DB 444 CGCAAGCGCGCCAGTGGAGCGCGCAGCGAGCGGAGCGCGGAGGAGCAGCAGATGATGA 503
QY 157 AlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAla 176
DB 504 GAGCATCGAAGACCTGGAGGATCTCAAGGCCACTCTGTGCGCGAGTGGTGGATGGC 563
QY 177 TyrValTrpGlyThrAsnValAsnThrProAspValLeuArgAlaIleArgArgPheLeu 196
DB 564 -----GGGCCCCGCGCTGGAGATCCACCACCGCTTCAAGAAGCTTCTCT-----GCG 608
QY 197 HisAsnTyrArg-SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleGluGln 216
DB 609 CACTCAGTCCGACAGCCACCGCCCAACAGTC-----TTCAAGGAGCGCATCAGCGA 659
QY 216 uThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAs 236
DB 660 CATGTGCAAGAGAACCGTGAGAGCTGTGTGTGTAAGTATGAGGACTTGGCAGCCAGGGA 719

```

```

QY 236 pProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspTh 256
DB 720 GCAGTGTCTGGCTTACTTCTGCTGAGGACACCGCGAGTGTGTCAGATCTTTGATGA 779
QY 256 xGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLys-----HisIle 273
DB 780 GGCTGCCCTGGAGTGTGCTGCGCATGTACCCCAAGTACGACCGCATCACCACCAT 839
QY 273 eGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAs 293
DB 840 CCATGTCCGATCTCCACCTGCTGCTGGTGGAGAGCTGCGCTCGCTGAGCAGCTGCA 899
QY 293 pIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGln 313
DB 900 TCTGAACAGCATGATCCGACACAGTGGGTGTGACACAGCTGCACTGGCTGCTGCCCA 959
QY 313 uIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProLeuValThrVa 333
DB 960 GCTCAGCATGTCAAGTACAACTGCAACAGTGCATTTCTGCTGGGTCTTCTTGCCA 1019
QY 333 lValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAs 353
DB 1020 GTCCAGAACACGAGGAGTGAAACACGAGCTCTGT-----CCTGAGTGCCAGTCGGCGG 1073
QY 353 nAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln 373
DB 1074 CCCCTTTGAGTCAACATGGAGGAGCACCATCTATCAGAACTATCAGAGCGTATCCGAATCCA 1133
QY 373 nGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTy 393
DB 1134 GGAGAGTCCAGGCAAGTGGCGGCTCGCGGCTGCCCGCTCCCAAGGAGCGCATCTCTCT 1193
QY 393 rAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
DB 1194 CGCAGATCTGTGTGACAGCTGCAACGCGAGGAGACGAGATAGAGTGTGCTAGGGGA 1253
QY 413 sAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAla-----LeuTyrIle 431
DB 1254 CAAC-----AACTATGATGCTCTCCCTCAACACTGCCAATGCTTCTCCCTCTTTCG 1304
QY 431 sThrTyrIleAspCysValHisValLysValSerAspArg-----GlyArgLeu 447
DB 1305 CACTGTCTATCTACCAACACAGTGGCGGAGAGGAGCAACAAAGTGTGCTAGGGGA 1364
QY 447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGln 467
DB 1365 GACCGATGAAGAT-----1377
QY 467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerIle 487
DB 1378 -----GTGAAGATGATCACTAGCTCTCCAA 1403
QY 487 sLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrGluLeuGln 507
DB 1404 GGATCAGCAGATCGAGAGAAGATCTTTCAGCATCTCTCTCCATCTATGTCATGA 1463
QY 507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysIleProSe 527
DB 1464 AGACATCAAGAGAGCCCTGCTGCGGCTCTGTCGGAGGGAGCCCAAAACCCAGGTGG 1523
QY 527 rGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerIle 547
DB 1524 CAAGCACAAGGTACGTGTGATATCAACGTCTCTTGTGCGGAGACCCCTGGCAGCA 1583
QY 547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
DB 1584 GTCCAGTTTCTCAAGTATATTGAAAGTGTCCAGCCGAGCCATCTTCCACACTTGCCA 1643
QY 567 gGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluTh 587
DB 1644 GGGGCGCTGCTGTGGCGGTACCGCGGTATGTCCAGCGGCAACCTGTGTCAGCGAGGTG 1703

```

```
QY 587 rValLeuGluSerGlyAlaLeuValuSerSerAspArgGlyIleCysCysIleAspGluPh 607
Db 1704 GACCTTGGAGGCTGGGGCCCTGGTCTGGCTACCGAGGAGTGTCTCATTTGATCAAT 1763
QY 607 eAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGluThrVa 627
Db 1764 TGACAGATGATGATGACGAGACAGACAGATCCATGAGCCATGAGGACACAGACAT 1823
QY 627 lSerValAlaLysGlyIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCy 647
Db 1824 CTCATCTCGAAGGTGGCATGCTACCTCCCTGCGAGGCTCGTGCACGCTCATTTGCTGC 1883
QY 647 sAlaLenProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGln 667
Db 1884 CGCCAAACCCCATAGGAGGCGGTACGACCCCTCGCTGACTTCTCTGAGAACGTGGACCT 1943
QY 667 uProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGl 687
Db 1944 CACAGAGCCCATCATCTACGCTTTGACATCTCTGTGTGGTGAGGACACCGTGGACCC 2003
QY 687 uGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVa 707
Db 2004 AGTCCAGGACGAGATCTGCGCGCTTCTGTGTGGCGACCCACGTCCAGACACCCAC 2063
QY 707 lSerLysGln----- 710
Db 2064 CAACAGGAGGAGGAGGCGGTGGCCAAATGGCAGCGCTGCTGAGCGCGCCATGCCAACAC 2123
QY 711 -----AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGl 727
Db 2124 GTATGCGGTGGAGCCCTGCCCGAGGAGGTCTGGAAGAAGTACATCATCTACGCCAAGGA 2183
QY 727 nHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGl 747
Db 2184 GAGGCTCCACCCGAGCTCAACGAGATGACGACGACGAGGTGGCCCAAGATGTACAGTGA 2243
QY 747 uMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProAr 767
Db 2244 CCTGAGGAAAGAACTATGCGCAGCAGCGCAGC-----ATCCCCATTACGGTGGC 2291
QY 767 gGlnLeuGluSerMet-IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluV 787
Db 2292 GCACATCGATGTCATGAGTC---ATGCGGAGGCGCCACGCGCGCATCTCTGCGGGACT 2348
QY 787 alValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db 2349 ATGTGATCGAAGACGACGTCAACATGCGCCATCCGCGTGATG 2389

RESULT 11
US-09-248-796A-4534
; Sequence 4534, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4534
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4534

Alignment Scores:
Pred. No.: 3,33e-59 Length: 1635
Score: 817.50 Matches: 188
```

```
Percent Similarity: 50.68% Conservative: 108
Best Local Similarity: 32.19% Mismatches: 181
Query Match: 17.74% Indels: 107
DB: 3 Gaps: 9
US-10-768-511-8 (1-901) x US-09-248-796A-4534 (1-1635)
QY 308 SerSerIleLeuProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSer 327
Db 1 ACCGGGGTTTTCCCAAAATGAAATATGTAATTTGATTTGTTGAAATGTGA-----54
QY 328 ProProLeuValThrValValLysGlyArgValGluGluProThrArgCysGluLysPro 347
Db 55 GTGGTTTGGTTTCCATACGTTCAAGATTCTAATCTGAAGTGAATAATCTCATCTCTGTACC 114
QY 348 GluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLys 367
Db 115 AATTGTCAAGTAAAGTTCCTTTTAAATTTGAATTCAGAGAAAATTTATATAGAACTAT 174
QY 368 GlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThr 387
Db 175 CAAGAATTACTTTTACAAGAGCTCCAGGTACTGTTCCAGCGGTAGATTGCCAAGACAT 234
QY 388 ValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGlu 407
Db 235 CGTGAAGTCATATTATTATCTGATTTAGTTGATTTGTCTAAACCCAGGTGAAGATATTGA 294
QY 408 ValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg 427
Db 295 GTTACTGTGTTTACAG-----AATAATTATGATCGGAATTAAAT 336
QY 428 Ala-----LeuTyrLysThrTyrIleAspCysValHisValLysLysSer 442
Db 337 GCCAAGAAATGAGTTCGCGGTGTTGCCCAATTTTGAAGCAATTCATTCATTAGAAGA---393
QY 443 AspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyr 462
Db 394 -----AAAGAAAGTAGTGCAATTT 411
QY 463 AlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeu 482
Db 412 ATGGGT-----GGTAATAATTAGTCAATATGTGGACAGAGAAGAAATACGTGAATTT 465
QY 483 LysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSer 502
Db 466 AGAAAAATTATCACAGAAAAAGGTATTATTGATAAAATCATTTGCCCTCAATGGCTCCATCC 525
QY 503 IleTyrGluLeuGluAspIleLysGlyLeuLeuCysGlnLeuPheGlyGlyCysAla 522
Db 526 ATATATGTCATAAAGATATTAAAACTGCCATTGCTTGTTCATTATTTCGTGTGTGCTT 585
QY 523 LysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAsp 542
Db 586 AAAGATGTTAATGGGAAACTTCCATTAGAGGTGATATTATATGACTATTATTAGGGAT 645
QY 543 ProGlyThrSerLysSerGlnLeuGlnTyrValHisLysIleAlaProArgIle 562
Db 646 CCAGGTACTGCTAAATCGCAATATTGAAATATGCTGAGAAAACCTCTAGTGTGTGCTGT 705
QY 563 TyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspPro 582
Db 706 TTTGCAACTGGTCAAGGTGCTCGGTGTTGTTGTTTACAGCATCGGTGTTAGAAAGACCCA 765
QY 583 GluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCys 602
Db 766 ATCACTCGTAATGGACATTAAGGTGGGCTTTGCTGCTTGTTCATTAAGAGGACTTGT 825
QY 603 CysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMet 622
Db 826 CTTATTGATGAATTTGATAAGATGATGATCAAGATAGAATCTTCTATTTCATGAAGCTATG 885
QY 623 GluGlnGlnThrValSerValAlaLysGlyIleIleAlaSerLeuAsnAlaArgThr 642
```


Db 922 TATGTCAGATATATCCAGATTGTCACACATGAATATATCAATTTACCAGCCAGCTTT 981
QY 671 uLeuSerArgpHe 675
Db 982 GTGGTCCCGCTTT 994

RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103398)..(103398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)

; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

```

```

Alignment Scores:
Pred. No.: 1.19e-40 Length: 1664976
Score: 656.00 Matches: 202
Percent Similarity: 46.75% Conservative: 143
Best Local Similarity: 27.37% Mismatches: 258
Query Match: 14.23% Indels: 135
DB: 3 Gaps: 26

```

US-10-768-511-8 (1-901) x US-08-916-421B-1 (1-1664976)

```

QY 177 TyrValTyrGly-ThrAsnValAsnIleProAsp-----ValLeuArgAlaIleAr 193
DB 10758 TTCGTATGGGAGTGTAGATATGGAATGAGATGAGACTTAATTTTATAGAGGAGTTAG 10699

QY 193 gArgPheLeuHisAsnTyrArgSerSerAlaHis----- 204
DB 10698 GGATTATTATTAACCTCTTTTAAAGAAATATTCATCAGGAAGATATTATCTAGACAATCA 10639

QY 205 -----AspLeuAsnSerLysTyr-----IleGlnI 213
DB 10638 AAGAGTTGTAGTTGATTTAAATCAACTACATTAACGGATTAAATGGAAATTTGTAGAAAT 10579

QY 213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTy 233
DB 10578 TTTTAAATAAATATCCAAAAAGGGATTTGATTTTATAAAGAGTGCTACAAACGATCTTA 10519

QY 233 xAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe 253
DB 10518 T-----TACACCTTAAGAATGATATCCAAACGTAATA----- 10482

QY 253 uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro---ThrPheGluLysHi 272
DB 10481 -----ATTGCTGTAAAAAATCTCCAAAAAATTTTAAAAACTAC 10444

```

```

QY 272 sIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSe 292
DB 10443 AAGAAAAGGGAAAAATTTTCAATAGAG-----GATATTAGGAGCAA 10402

QY 292 rAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIlePr 312
DB 10401 AACTTTTAGAAAAATTTAGTTGAATTTGAAGGAATCATTTGTTATGGCATCAAAATAGACC 10342

QY 312 oGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValTh 332
DB 10341 AATGCTTAAAAAGCTTATTATATATGTCAAAAATGTGGAAGAGAA----- 10296

QY 332 rValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaAr 352
DB 10295 -GTTGTTAGAAAAATAGATATCTTAAACACTGACTCTGAAAAAGCTGTTTGTGAATGTGG 10237

QY 352 gAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLue 372
DB 10236 AGCTGAGTTAAACCTAATTGAGCATAAATATATATATGACTTCCAGAGATTAAAGT 10177

QY 372 uGlnGluThrProAspAlaIle-----ProGluGlyGluThrProHisThrValSerMe 390
DB 10176 TCACAGCCCATTTGGATTTAATGGAAAAATCCTGAA--GAGCCGCCAAAAATACATACTGT 10120

QY 390 tCysLeuTyrAsnThrMetValAspAlaValLysProGly-----AspArgIleGl 407
DB 10119 GTTCTTTAGAAAAACAGT-----CCTGGAATATATGCTGGAAGGGTGAA 10078

QY 407 uValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuAr 427
DB 10077 GATAACTGCGATC-----CCATATAAGTTAAGAGAGTAAAAAGCTA----- 10035

QY 427 gAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLue 447
DB 10034 -CCAATCTATGACATACATGTTAAAGCCCTACACTGTGAGGTTTGGAT----- 9987

QY 447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGl 467
DB 9987 ----- 9987

QY 467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerly 487
DB 9986 -GGGGAAGTAAAAATTTAAATTAACAAATTCAGATATTGAAATATATTTAAAAAATAGCTAA 9928

QY 487 sLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrGluLeuGl 507
DB 9927 AAGAAAAGATGTTGTTAATATACTTTCAGATAGATTAAATCCAGAGATTAAAGGGGATTC 9868

QY 507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSe 527
DB 9867 TGCAATTAAGAAGGGCTGCTTACTACACAGATAAAGGAGTTAAAAA----- 9819

QY 527 rGlyAlaSerPheArgGlyAspIleAsnValLeuValGlyAspProGlyThrSerly 547
DB 9818 ----CCTGGAAGAGGGCTGATATTATATATTTAATCAGACCCAGAGATTGGAAA 9763

QY 547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyAr 567
DB 9762 AACAGTCATACTAAGAAGAGATTGCTGAATC---CCAGAAATTTATACGGTCTCTTTAC 9706

QY 567 ggLySerSerAlaValGlyLeuThrAlaTyrValThrLysAsp---ProGluThrArgGl 586
DB 9705 TACCGCCACTGCTGTTGGCTAACTGCGCTGTTGTTGGGAGAGACAGAGATTGGTGA 9646

QY 586 uThr-----ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIl 604
DB 9645 AGATACATGGTTATTAAAGCAGGTTTGTAGTTAAAGCCCAAGAGAACTGCATGTAT 9586

QY 604 eAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGl 624
DB 9585 AGATGAGCTA---ACTGTTTAAAGAACTGCAGACTATGTTTGGAGGCTATGGAGAG 9529

QY 624 nGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVa 644

```



```
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
```

```
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
```

Alignment Scores:

Pred. No.:	1.19e-40	Length:	1664976
Score:	656.00	Matches:	202
Percent Similarity:	46.75%	Conservative:	143
Best Local Similarity:	27.37%	Mismatches:	258
Query Match:	14.23%	Indels:	135
DB:	3	Gaps:	26

US-10-768-511-8 (1-901) x US-09-692-570-1 (1-1664976)

QY	177	TyrValTrpGly-ThrAsnValAsnIleProasp	-----ValLeuArgAlaIleAr	193
DB	10758	TTCTGATGGGATGTAGATATGGAATTGAGAGATGAAGACTTAATTTTATAGAGGAAGTTAG	10699	
QY	193	gArgPheLeuHisAsnTyrArgSerSerAlaHis	-----	204
DB	10698	GGATTATTAACTGCTTATTAAAGAAATATTTCATCAGGAAGATATTATCTTACACAATGA	10639	
QY	205	-----AspLeuAsnSerLysTyr	-----IleGlnIle	213
DB	10638	AAGAGTTGTAGTTGATTTAAATCAACTCTACAATTACGGATTAATGGAATTTTGTAGAATT	10579	
QY	213	eileGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr	233	
DB	10578	TTTAATAAATAATCCACAAAAGGGATTTGATTTTATAAAGAGTGCTTACACAGATGCTTA	10519	
QY	233	xAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe	253	
DB	10518	T-----TACACCTTAAGAAATGAATATCCACAAACGTAATA-----	10482	
QY	253	uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro---ThrPheGluLysHi	272	
DB	10481	-----ATTGCTGTAAAAAATCTCCCAAAAAATTTTAAAAACTAC	10444	
QY	272	slleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSe	292	
DB	10443	AAGAAAAGGGAAATTTTTCACATAGAG-----GATATTAGGAGCAA	10402	

US-10-768-511-8 (1-901) x US-09-828-062-5 (1-4348)

QY 1 MetGluAenAenAspAlaIeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
DB 1554 ATGGAAATAAATGATGSCATTGACATTTGGAGCGGTGCTGCCATATCCTTCGCAATCT 1613
QY 21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaIaSer 40
DB 1614 GAGAGAGTGTCTACGCCATTGCCCAAGTAACATCACCAGAGCTTCGACAAATGCAGCCTCA 1673
QY 41 ProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArgArgGly 60
DB 1674 CCCGTGCGCGGCGAGGCGGCTACCGCAGACCCCTACATCTGCAGTTCCGAGGAGAGGG 1733
QY 61 ArgGluThrAspSerAlaArgArgArgSerArgSerLeuGlyAsnSerVal 80
DB 1734 AGAGAAACGGATTCGGCTCGTGGTAGGAGAGTCGATCTCGCAGTTAGGCAATTCCTGTT 1793
QY 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
DB 1794 TATAGTTCCTCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
DB 1854 GTTTAGCGTACCCCACTCGGTACACTATGGGTACCCCATCGTTCCATCGTGGCAGGCCA 1913
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
DB 1914 CAGTACAAACAGCGCAGTGTGGTTCCAGGGGAAAGCCCTTACATCGGAGACGTCTGA 1973
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
DB 1974 TCTCAATCCAGAAACCCGGGCATCGATCTCCTCAAGGGAACCTAGTGTGATGGCGGT 2033
QY 161 ProSerGluSerAlaGluProAspThrLeuGlyGlyGluTyrAlaTyrValTyrGly 180
DB 2034 CCTCTGATCTGCTGAGCCAGATGACACTTTGGTGGAGAAATATGCTTATGTTGGGG 2093
QY 181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
DB 2094 ACGAATGTTAACTTCCAGATGTGCTTAGGGCGATTCGTCTCAACAAATATTCGT 2153
QY 201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluLeuThrValGluArg 220
DB 2154 TCGAGTGTCTGATGATCTTAATTCCAAGTACATCCAGATCATAGGAGACCTGTGGAGCGT 2213
QY 221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
DB 2214 GAGGAGGATACTTAAATATCGACATGTGACACATTTATGACATCTCTGATCTATAC 2273
QY 241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
DB 2274 GCAAAAATTTGTTGATACCCACTCGACATCATCCCTCCCTGTTGGACACATGAGTGTGAGAA 2333
QY 261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
DB 2334 GTTGCTACTCTTTACTACCAACCTTTGAGAGCATATTTGAGCCCGACACCTTTCAATCTC 2393
QY 281 LysAlaSerValHisMetArgGluLeuAenProSerAspIleTyrAspLeuValSerVal 300
DB 2394 AAAGCATCGGTGCATCGGTGAATCAACCCCTTCAGATATAGACAAATTTGGTTCTGTT 2453
QY 301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
DB 2454 AAAGGAATGGTTATCCGGTGCAGTCTATCATCTGAAATTTAGGGGGCCTTCTCAAA 2513
QY 321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
DB 2514 TGTTTAGTGTGGTGCATCGGCTCCGCTAGTGTACAGTTGTTTAAAGGGCGGGTTCAGGAG 2573
QY 341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaValSerLeuIleHisAsn 360
DB 3654 CACTATGAAACTATGAATTTCAAAGCAGGACGCGCTTAGATCTACAAACACTTACCGCG 3713

DB 2574 CCAACAAGGTGTGAAAGCCAGAAATGTGCAGCAGGAATGTATGTCTCTTTATTCACAAT 2633
QY 361 ArgCysThrPheAlaAenLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
DB 2634 CGATGCATCTTTTGGAAAATAGCAGATAGTGGGTCTTCAAGAAACTCCAGATGCCATTCCT 2693
QY 381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
DB 2694 GAAGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGTGTTGATGCTGTG 2753
QY 401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
DB 2754 AAGCTGGAGATCGTATTGAGGTAACAGGAGTTCCTCAAGGCCATCGGCAGTTCCAGTTGCT 2813
QY 421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
DB 2814 CCGAATCAACGAACATTTACGAGCATTTGTATAGACCTTACATCGATTGGTGCACGTCAAG 2873
QY 441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAenAsp 460
DB 2874 AAGTCTGACGGGTGCACTGCAAACTGAAGATCCTATGGAGATGGATAGGAGAAATGAT 2933
QY 461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
DB 2934 ATGTATGCTGGTATCATGAAGTGATACTTTCAGAGCTGCTATGAAGCAAGATTCAA 2993
QY 481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
DB 2994 AAACCTTAAAGAGCTGTCCAGCTCCCGGCAATTTATGATAGACTTTTCAAGGTGCTGGCT 3053
QY 501 ProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
DB 3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113
QY 521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
DB 3114 AAGGCTAAGAAAATTCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3173
QY 541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
DB 3174 GGGGACCTGTGTACAGTAATCTCAGCTGCTTCAAGTGTGTCAGCAAGATAGCTCCTCGT 3233
QY 561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
DB 3234 GGAATCTACACTAGTGGCGAGGAAGTTCCGGCGTTGGGCTGACAGCGTATGTAACAGAG 3293
QY 581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
DB 3294 GATCCAGAACTCCGAGAGCGGTATTCGAGAGCGGAGCTTTGGTTCTTTAGTATCGTGGG 3353
QY 601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
DB 3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCGCGAAGCATGCTTTCATGAG 3413
QY 621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
DB 3414 GTAATGGAGCAACAAACGATCTGTAGCCAAAGGGGTATCATTTGCTTCGTGAAACGCT 3473
QY 641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
DB 3474 CGGAGCTGTCTCTTCATGTGCAAAATCCTAGTGGTCCCCGATACAATGCGCGCTTCTCT 3533
QY 661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
DB 3534 GTGATTTGATAACATCCAGCTTCTCCCACTCTACTTCTTAGATTTGATTTAATTTACTTFA 3593
QY 681 MetLeuAspLysProAspGluAenAspArgArgLeuAlaArgHisLeuValAlaLeu 700
DB 3594 ATGCTCGCAAAACAGCAGCAGCAAAACATCGTCTGCTCGCCAGGCACTCTCGTGGCTTFA 3653
QY 701 HisTyrGluAenTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
DB 3654 CACTATGAAACTATGAATTTCAAAGCAGGACGCGCTTAGATCTACAAACACTTACCGCG 3713

QY 721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
DB 3714 TATATCACTATGCTCGTCAGCATGTACATCTACATTAAGTGAAGCTGCTGAAGAT 3773
QY 741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
DB 3774 TTGATTAATGCTATGTTGAGATGCGCAAGGGCAACTTTCTCTGGGAAGCAGTAAAAAG 3833
QY 761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
DB 3834 GTGATACAGCCACACCTCGGCAACTCGAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893
QY 781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
DB 3894 CGAATGAGATTTCTGAAGTGTGAGAGAAAGTTGATGACGACAGAGCTGTGCGCCTTTTA 3953
QY 801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
DB 3954 GACGTCGCTTTGACGAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013
QY 821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuAlaAlaLeu 840
DB 4014 ATCAGCTGAGAGTGTGCGCAGCGAGCGTATTCGTGCGGCCCACTTGTAGCTGCTCTG 4073
QY 841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
DB 4074 CGAGAGCTTATAGCAGATATAATTTCACTGCGCAGCTCTCTGGCTTGAAGCAGTACAG 4133
QY 861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
DB 4134 CTTCTTGAGGATATCCGAGGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAATAAT 4193
QY 881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
DB 4194 GCTCTGGGTAGCTCCAGGAGAGGCTTTCTTACTGTCTCCTGATGATAGTCAAGAGA 4253
QY 901 Val 901
DB 4254 GTT 4256

RESULT 2

US-10-768-511-5
; Sequence 5, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/10/768,511
; CURRENT FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4348
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-10-768-511-5
Alignment Scores:
Pred. No.: 0 Length: 4348
Score: 4609.00 Matches: 901
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-768-511-8 (1-901) x US-10-768-511-5 (1-4348)
QY 1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
DB 1554 ATGGAAATATGATGACATTCGATTCGAGCCGTCGTCCCATATCTCTCGCAATCT 1613
QY 21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAsnAsnAlaAsn 40
DB 1614 GAAGAGAGTGTACGCCATTCGCCGCACTAACATCACCAGAGCTTCGACATGCACCTCA 1673
QY 41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgGly 60
DB 1674 CCGTGGCCGCGGAGGCGGTACGCGCAGACCCCTACATCTCGAGTTTCGAAGAGAGGG 1733
QY 61 ArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
DB 1734 AGAGNAACGATTCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATCTGTT 1793
QY 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
DB 1794 TATAGTTCCCTTACGATGCGGGACTCTCGAACTCTCGAACTCCAGTGGCTACTCCG 1853
QY 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
DB 1854 GTTTACGCTACCCCGAGTCGTCACCTATGGGTACCCCATCGTCCATCGTGGCAGCGCA 1913
QY 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
DB 1914 CAGTACAAACAGCGCAGTGTGGTTCCAGGGGAGAGCCCTCTACATCGGAGAGCTCGA 1973
QY 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
DB 1974 TCTCAATCCAGAAACCCCGGCATCGATCTCTTCAAGGAACCTAGTGTCTGATGGCGT 2033
QY 161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGlyTyrAlaTyrValTyrGly 180
DB 2034 CCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGATATGCTTATGTTGGGGG 2093
QY 181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
DB 2094 ACGAATGTTAACTTCCAGATGTGTTAGGGCGATTCGTGATTTCTCCACATTAATCGT 2153
QY 201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
DB 2154 TCGAGTGTCTCATGATCTTAATTCAGTACATCCAGATCATAGAGAGAGACTGTGGAGCGT 2213
QY 221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
DB 2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCTCTGATCTATAC 2273
QY 241 AlaTyrIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
DB 2274 GCAAAAAATGTTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTGAGAA 2333
QY 261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
DB 2334 GTTGTACTCTCTTTACTACCAAGTTTGAGAAGCATATTAGAGCCAGACCTTTCAATCTC 2393
QY 281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
DB 2394 AAAGCATCGGTGCACATGCGTGAACCTCAACCCCTCAGATATAGACAAATTTGTTCTGTT 2453
QY 301 LysGlyMetValIleArgCysSerSerIleIleProGluIleIleGlyAlaPhePheLys 320
DB 2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATCTGAAATTAAGGGGGCCCTCTTTCAA 2513
QY 321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
DB 2514 TGTTTAGTGTGTGTGCTACCTCGCTCGCTAGTTACAGTTGTTAAAGGGGGGTTGAGGAG 2573
QY 341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
DB |||||

```
Db 2574 CCAACAAGGTGTGAAAGCCAGATGTGACGCGAATGCTATGTCTCTATTACCAAT 2633
Qy 361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db 2634 CGATGCACCTTTGCAAAATAGCAGATAGTGGTCTTCAAGAAACTCCAGATGCCATTCCT 2693
Qy 381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db 2694 GAAGAGAGAGCTCCACACACAGTCAGCATGTGTTTATACAACATATATGTTGATGCTGTG 2753
Qy 401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db 2754 AAGCTGGAGATCGTATTGAGGTAAACAGAGTTTCAAGGCCATGGCAGTTCCGATTGGT 2813
Qy 421 ProLenGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db 2814 CCGAATCAACGAACATTACGAGCATGTATTAAGACCTTACATCCATTTGGTGCACGCTCAAG 2873
Qy 441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db 2874 AAGCTGACAGGGGTGCGACTCGAACTGAAGATCCTATGGAGFTGGATTAAGGAGATGAT 2933
Qy 461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db 2934 ATGTATGCTGGGTATCATGAAGTGATACTTCAGNAGCTGCTPATGAAGCAAAAGATTCAA 2993
Qy 481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgIleuSerArgSerIleuAla 500
Db 2994 AAATTAAGAGCTGTCTCAAGCTCCGGGCACTTTATGATAGACTTTTCAAGTCCGTGGCT 3053
Qy 501 ProSerIleThrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db 3054 CCAAGCATTTGGGAGCTTGAAGATATTAAGAAGGGTCTTCTTGGCAGCTCTTTGGTGGG 3113
Qy 521 LysAlaLysLysIleProSerGlyValaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db 3114 AAGGCTAAGAAATTCATCTGGAGCATCTTCCGAGGTGACATCAATGTTTTACTTGT 3173
Qy 541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db 3174 GGGGACCTGTGTACCAGTAAATCTCAGCTGCTTCAAGTATGTGCACAAGATAGCTCCTCGT 3233
Qy 561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db 3234 GGAATCTACATAGTGGGCGAGGAAGTTCCGCGGTTGGGCTGACAGCGTATGTAAACGAAG 3293
Qy 581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db 3294 GATCCAGAACTCCGAGACCGGTATTGGAGAGCGAGGCTTTGGTTCTTAGTGATCGTGGG 3353
Qy 601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db 3354 ATATGCTGTATCGATGAGTTCCGACAAATGTCTGATAATGCCCAAGCATGCTTCATCAG 3413
Qy 621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerIleuAsnAla 640
Db 3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTTGCCCTCGCTGAACGCT 3473
Qy 641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db 3474 CGAGCTCTGCTTGTGATGTGCAATCTTAGTGGGTCCGATACAAATGCCGCGCTTTCT 3533
Qy 661 ValIleAspAsnIleGlnLeuProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db 3534 GTGATTGATTAACATCCAGCTTCTCCAACTCTACTTCTTAGATTGATTTAATTTACTTA 3593
Qy 681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db 3594 ATGCTCGAACAACCCAGACGAGCAAAACGATCGTCTCGCCAGGCATCTCGTGGCTTTA 3653
Qy 701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
Db 3654 CACTATGAACATATGAAGTTTCAAGCAGGACGGCTTAGATCTTACAAACACTTACCGG 3713
```

```
Qy 721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
Db 3714 TATATCACCTATGCTCGCTCAGCATGTATCATCTCATATTAAGTATGAAGCTGCTGAAGAT 3773
Qy 741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
Db 3774 TTGATTAAATGGCTATGTGAGATGCGCCAAAGGGCAACTTTCCTGGNAGCAGTAAANAAG 3833
Qy 761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
Db 3834 GTGATAACAGCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893
Qy 781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db 3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAAGCTGTGCGCTTTTA 3953
Qy 801 AspValAlaLeuGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
Db 3954 GACGTCCCTTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013
Qy 821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaLeu 840
Db 4014 ATCAGCACTGGAGTGTCCGCCAGCGAGCGTATTCTGTCGGCCAACTTCTAGTCTCTG 4073
Qy 841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
Db 4074 CGAGAGCTTATAGCAGATAAAATTTTCCCTGGCAGCTCCTCTGCTTGAAGCAGTCTAG 4133
Qy 861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
Db 4134 CTCTTCAGAGGATATCCCGAGCAAGCAGTGTGACGCTTAGTTTGCAGGATATTAAAAAT 4193
Qy 881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
Db 4194 GCTCTGGGTAGCTCCCAAGGAGGAGGCTTCTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253
Qy 901 Val 901
Db 4254 GTT 4256
```

RESULT 3
US-10-425-114-28735
; Sequence 28735, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 28735
; LENGTH: 2841
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4763-010-C6_FLI
US-10-425-114-28735

Alignment Scores: 8.18e-209 Length: 2841
Pred. No.: 2550.00 Matches: 514
Score: 73.24% Conservative: 143
Percent Similarity: 57.30% Mismatches: 176
Best Local Similarity: 55.33% Indels: 64
Query Match: 7 Gaps: 17
DB:

[illegible]


```
QY 725 AlaArgGlnHisValHisProThrLeuSerSerAspGluAlaAlaGluAspLeuLeuLeuGly 744
D 2116 GCAAGGAAGTATATTATACCAACAGTTATCTGATGAAGTCGAGAGAGTTAACTCGTGGC 2175
QY 745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
D 2176 TATGTCGAGATGAGAAAAGAGGGAATAGCCCTGGGAGCAGNAGAGGTATACAGCA 2235
QY 765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
D 2236 ACCGCTAGACAAATAGAGAGTTTGATCGTCTCAGCGAAGCAATAGCCCGGAATGCGGTTC 2295
QY 785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
D 2296 TCTGAAGTGGTCGAGGTGCGGATGTTGTGAGGCAATTCAGGCTTCTTTGAAGTCGCCATG 2355
QY 805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrGly 824
D 2356 CAGCAGTCTGCAGCGGATCATGCAACTGTGTAGATTGATATGATCTCATGACGGGG 2415
QY 825 ValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaLeuAlaGluLeuIle 844
D 2416 ATATCCGCAAGCAAGGAGGAGGCGGAGAACCTCGTTCGCGCAACCCGTAACCTGATT 2475
QY 845 AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAsp 864
D 2476 GCGGAGAAAATGACAGCTTGGAGGCCCTCG---ATGCGCATGATTGAGTTGCTGGAGGAA 2532
QY 865 IleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySer 884
D 2533 CTGAGGAAGCAGAGCTCATGGAATTCATATGCAAGAACTCCGCGGTGCTCTTGGCAC 2592
QY 885 LeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
D 2593 CTGATGACTGAAGGCGGTGGTTATTCATGAGAGCAACGTGAGGAGAGTT 2643
```

RESULT 4

```
US-10-437-963-44639
; Sequence 44639, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 44639
; LENGTH: 3054
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MKT4530_47680C.1
US-10-437-963-44639
```

```
Alignment Scores:
Pred. No.: 7, 06e-203 Length: 3054
Score: 2481.50 Matches: 512
Percent Similarity: 71.96% Conservative: 132
Best Local Similarity: 57.21% Mismatches: 180
Query Match: 53.84% Indels: 73
DB: 7 Gaps: 15
US-10-768-511-8 (1-901) x US-10-437-963-44639 (1-3054)
```

```
QY 24 SerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerProValAla 43
D 37 AATTCCCTCTCCCTCCGTCGTCGTCGCGGATGT---CGACCGTCGAGCCCTCTCCCG 93
QY 44 GlyArgAlaValArgGlnThrProThrSerAlaValArgArgGlyArgGluThr 63
D 94 -----GGACCAACTCTCTCCCGCTCAGTCGGGGCGCGCGCGG----- 135
QY 64 AspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSer 83
D 136 -----GGGGCGCGCGCGGTTC-----GCCAGC 162
QY 84 ProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
D 163 CCGTACCCGTCG---TCCCGCTCCCTCGCGGG-----TTCAG 198
QY 104 ThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLys 123
D 199 ACGCCGCGCACCGCGGGCGCGCCCTTC-----GGCGGCGCGCGCGCGGCGG 252
QY 124 GlnArgSerGluLeuGlySerGlnGlyLysPro-----LeuHis 136
D 253 CAGCGCA-GAATGGAGCGGGCGGCTTCCCGCGACCCCGTCCACCCCATGTCCAC 311
QY 137 ArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
D 312 CGACGAGTCC-CGCTCTCTCGGAGCGCGGGACGAGACACCCCGGAGCGCG 370
QY 157 Ala--AspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyr 176
D 371 GCGCGCGCGCGCGCGCGCGCGCGCGCGCG-----G 406
QY 176 IatYrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheL 196
D 407 TGTTCTGTCGGGCAACCAATCAGCGTGCAGGAGCTCAACCGCCATCTCTCCGTTCC 466
QY 196 euHisAsnTyrArgSerSerAlaHisAspLeu-----AsnS 208
D 467 TCCGCCACTTCCG---GACCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGAGG 523
QY 208 erLysTyrIleGlnIleGluThrValGluArgGlu---GluAspThrLeuAsnI 227
D 524 GCAAGTACATCGCGCCATCCACCGCATCTCTCGAGCTCGAGGGCGGGGAGTCTCGACG 583
QY 227 leAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrP 247
D 584 TCAACGCCACGAGCTGTTCGACCCAGCACCGCGGACCTCTACGGCAAGATGTCAGGTACC 643
QY 247 roLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuP 267
D 644 CGCTCGAGGTGCTCGCCATCTTTGACATCTGTGCTCATGAGCTCGTCGCGCGCATCGAGC 703
QY 267 roThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetA 287
D 704 CCTCTTCGAGAGACATCCAGACGAGATCTCAACCTCAAGTCTCGGTTGCTTGA 763
QY 287 rgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgC 307
D 764 GGAATCTCAACCTTCTGATATCGAGAAGATGGTGTCCATCAAGGGTATGATAATTCGT 823
QY 307 ysSerSerIleIleProGluLysGlyAlaPhePheLysCysLeuValCysGlyHisS 327
D 824 GCAGCTCGGTGATTCAGAGCTCAAGGAGGCTGTTCCTCGCTGCTGTGTTGCGGGTCT 883
QY 327 erProProLeuValThrValValLysGlyArgValGluProThrArgCysGluLysP 347
D 884 ACTCTGAGCTGTATGTTGATCGAGGAGGGTTACTGAGCGCGCACATCTGTTCAGAAAG 943
QY 347 roGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnL 367
D 944 AACAAATGTAAAGCTACAAACTCTATGACTCTTGTCATAACCGATGCGAGGTTTCAGATA 1003
```



```
QY 367 ysGlnIleValargLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisT 387
Db 1004 AGCAGATCATAAAGTTGAGGAAACACACAGATGAGATACCAAGAGGTGGCACTCCACATA 1063
QY 387 hrValSerMetCysLeuTyrAsnThrMetValAspAlaValysProGlyAspArgIleG 407
Db 1064 CTGTGTAGTGTCTTGATGACAGATAAGCTTGTAAGTCTGGAAGCCTGGAGACAGGGTTG 1123
QY 407 luValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuA 427
Db 1124 AGATAACTGGGATATACAGAGCCATGAGTATTAGAGTTGGCCCACTCAGAGGACAGTGA 1183
QY 427 rgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgL 447
Db 1184 AGTCGATTTC-----AAGAAGACACAGCAAGTCTAGAC 1216
QY 447 euGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisG 467
Db 1217 TTCTATGTTGAGGACTCCATGGAAACTGATAACCCCAAT-----GCTA 1258
QY 467 luSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerL 487
Db 1259 ACAACAACTGAGATGATGTTCTCAGAGATAAGTTGAGAAATTAAGAAGTTGTCA 1318
QY 487 ysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrGluLeuG 507
Db 1319 AGTTCGCCAGATATATACAGATTAACCTAGGTCAATTGGCTCCAAACATATGGGAGCTGG 1378
QY 507 luAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProS 527
Db 1379 ACAGAGCTTAAAGAGGGGCCCTCTTGGCAGCTTTTGGTGGAAATGCTTTGAGGCTTCCT 1438
QY 527 erGlyAlaSerPheArgGlyAspIleAsnValLeuValGlyAspProGlyThrSerL 547
Db 1439 CTGGAGCTAGTTTCCAGGCGACATCATATTTTCTGTTGGTGGAAATGCTTTGAGGCTTCCT 1498
QY 547 ysSerGlnLeuLeuIntYrValHisLysIleAlaProArgGlyIleTyrThrSerGlyA 567
Db 1499 NATCCAGCTTCTCCAATACATGACAAACTGCTCTCTGTCGTGGCATTTATACAGTGGCA 1558
QY 567 rgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluT 587
Db 1559 GAGGAAGTTTCAGCTGTGGCCCTTACTGCATAGCTTACCAAGGATCCTGAAACTGGTGGAA 1618
QY 587 hrValLeuLysSerGlyAlaLeuValLeuSerAspArgGlyLysCysLysIleAspGluP 607
Db 1619 CTGTCTTCGAGAGTGAGCGCTGTTTTCGAGTGACAAAGGTGTGTGTATTATGATGAAT 1678
QY 607 heAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrV 627
Db 1679 TTGATPAAGATGCTGATATGCTCGAAGCATGTTACATGAGGTGATGGAACAACAGACTG 1738
QY 627 alSerValAlaLysGlyLysIleAlaSerLeuAsnAlaArgThrSerValLeuAlaC 647
Db 1739 TCTCCATTGCCAAGGCTGGAATAATGTCATCTTTAAATGCCAGAACATCAGTTCTAGCAT 1798
QY 647 ysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnL 667
Db 1799 GTGCAAAATCCTATGAAATCACGTTATAATCCAAAGGCTCTCTGTGATTGCAATATCCATC 1858
QY 667 euProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspG 687
Db 1859 TTCTCTCAACACTGCTTTCTAGTTTGGACTCATTTATCTGATATTGGACAGGCGAGATG 1918
QY 687 luGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluV 707
Db 1919 AGCAAACTGATAGAGCCCTGGCTTAAGCATATTGTTTCGTTGCATTTTGGAGATCCAAACA 1978
QY 707 alSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgG 727
Db 1979 TAGAGGAGCTCGAGGCTTGGATTTCGCAACACTTGTAGCTTACATTAAGTTATGCAAGGA 2038
QY 727 lnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValG 747
```

```
Db 2039 AGCATATACAAACCAAGTTATCTGATGAAGCTGCGAAGAATTGACTCCGCGCTATGTG 2098
QY 747 luMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProA 767
Db 2099 AGATGAGAAAGAGGAACACGCCCTGTTAGCAGAAAGAGGTCTAATCTGCGACAGCTC 2158
QY 767 rgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluV 787
Db 2159 GACAAATTCAGAGCTTGATTCGGCTCAGTGAAGCACTGCGCCGAATTCGATTTCTCTGAA 2218
QY 787 alValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnS 807
Db 2219 TGGTTGAAGTACAAGATGTTAGAGGCTTTCAGGCTTCTCGAAGTTGCCATGCGACAA 2278
QY 807 erAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerA 827
Db 2279 CGGCAACTGATCATGCCACTGGTACATCGATATGATATCATGACTGGATATCTG 2338
QY 827 laSerGluArgIleArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspL 847
Db 2339 CGAGCGAAAGGCGAGAGCGGACAAATCTCGTTGCAGCAACCGCAACCTTGTGATGAGA 2398
QY 847 ysIleSerProGlySerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgS 867
Db 2399 AAATCGAGCTTGGAGGGGCCCTCA---GTCCGAATGATTGAGTTGCTGGAGAAATTAGGA 2455
QY 867 erGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnG 887
Db 2456 AGCAGAGCTCTATGGAAGTTCTATCTGCATGATCTTCGCGGTGCTCTTGGCATCTGATCA 2515
QY 887 lyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
Db 2516 CAGAGGGTGGCTAGTCTATCATCCATGAGACAGCGTCAAGAGG 2556
```

RESULT 5

```
US-10-425-115-98898
; Sequence 98898, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98898
; LENGTH: 2186
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2186)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRF4577_21708C.1
US-10-425-115-98898
```

```
Alignment Scores:
Pred. No.: 1,67e-158 Length: 2186
Score: 1963.50 Matches: 381
Percent Similarity: 81.79% Conservative: 86
Best Local Similarity: 66.73% Mismatches: 81
Query Match: 42.60% Indels: 24
DB: 8 Gaps: 3
```

US-10-768-511-8 (1-901) x US-10-425-115-98898 (1-2186)

QY 331 ValThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAla 350

Db 11 GTTCATGGTTGATAGAGGAGTAACTGAAACCGACATTTCTCAGAAAGACAAATGTAAA 70
Qy 351 AlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleVal 370
Db 71 GCCACGAATTTCTATGACCCCTAGTCACAAACCGATGTCAGATTTTTCAGACAAGCAGATCAT 130
Qy 371 ArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMet 390
Db 131 AAGTTGCGAGAAACACACAGACGAGATACCAAGAGTGCCACTCCACATACATAGTTAGTGTC 190
Qy 391 CysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGly 410
Db 191 TTGATGCAATGATAGCTTTGTTGATGCTGGAAGCCCTGGAGATAGGGTTGAGATTAACCTGGA 250
Qy 411 ValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyr 430
Db 251 ATATACAGAGCTATGAGTATTGCAATTGGAATTCGACCAAC----- 285
Qy 431 LysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGlu 450
Db 286 -----ATAAAGAGACAGACAAGAGTCTAGGCTTCATGTGGAG 321
Qy 451 AspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThr 470
Db 322 GACACCATGGATATTGATAATTCTTAAC-----GCTAGCAAAATCTACT 363
Qy 471 SerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGly 490
Db 364 GAAGAGGATTTTCTTAGTGATAAGGTTGAGAAACTAAAGAGCTTTCGAAGTTGCCGTAT 423
Qy 491 IleTyrAspArgLeuSerArgSerLeuAlaProSerIleThrGluLeuGluAspIleLys 510
Db 424 ATCTATGAAGATTGACTAGATCATTTAGTCTCAACATATGGAGTGGATGATGTCAAA 483
Qy 511 LysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyValaSer 530
Db 484 AGAGTCTCTTTGCGAGCTTTTTCGGCGCAATCCCTTGAAGCTTCTTCTGGAGCTAGT 543
Qy 531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550
Db 544 TTCGGGGTGATCATCAATATTACTTTGCGGGGACCCCTGGAAACAAGTAAATCCAGCTT 603
Qy 551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570
Db 604 CTCACGATCATGCATAAACTGCTCTCTGCTGATCTATACGATGAGTGGTAGAGGAGTTCT 663
Qy 571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
Db 664 GCTGTGTGCTTACTGCTTATGTTACCAAGACCCCTGAGACTGCGCAAACTGTTCTAGAA 723
Qy 591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610
Db 724 AGTGGGACACTGTTTTCAGTGACAAAGGTGTTGTTGTCATAGATGAGTTGATAAGATG 783
Qy 611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValala 630
Db 784 TCTGATAATGCCGGAAGATGTTACACGAGGTGATGGAACACGACGACAGATATCCATTGCG 843
Qy 631 LysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnPro 650
Db 844 AAGCGTGAATAATTGCATCTTTAAACGCTAGGACATCTGCTCTGGCATGTGCCAATCCT 903
Qy 651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThr 670
Db 904 ACTGAATCACGTTACAAATCCAGGCTCTCTGTAATTGACACATCCACTTAGCGGCAACG 963
Qy 671 LeuLeuSerArgPheAspLeuIleTyrLeuMetMetLeuAspLysProAspGluGlnAsnAsp 690
Db 964 CTACTTTCAAGATTGACCTGATTTATCTTCTTGGACAAGGCGGATGAGCAAACTGAT 1023
Qy 691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710

Db 1024 AGGCGCCTGGCAAGCATATTTGTTGCTGTGCAATTTTGAATCCAAATTTTAGAGGAGCTC 1083
Qy 711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
Db 1084 GAGTCTTGGACTTGGACACACTAGTTTCTTACATAAAGCTATGCAAGGAAGTATATTTCAG 1143
Qy 731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
Db 1144 CCACAGTTATCTGATGAAGCTGCAAGAGTTAACTCGTGGCTATGTGCGATGAGAA 1203
Qy 751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
Db 1204 AGAGGGAATAGCCCTGGAGCAGAAAGAGTCTATACAGCAACCGCTAGACAAATAGAG 1263
Qy 771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
Db 1264 AGTTTGATCCGCTCTCAGCGAAGCATTAGCCCAATGCGGTTCTCTGAAGTGGTCGAGGTG 1323
Qy 791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db 1324 CGGATGTTTGGAGGCATTCAGGCTTCTTGAAGTCGCGCATGCGAGCTCTGCGACGGAT 1383
Qy 811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArg 830
Db 1384 CATGCAACTGGAGCATTTGATATGATCTGATCATGACGGGGATATCGCAGCGAAAGG 1443
Qy 831 IleArgAlaAlaAsnLeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSerPro 850
Db 1444 CAGAGCGGGAGAACCTCGTTGCGCAACCCGTAACCTGATTGCGGAGAAAATGCAGCTT 1503
Qy 851 GlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSer 870
Db 1504 GGAGCGCCCTCG---ATGCGCATGATTGAGTTGCTGGAGGAACCTGAGGAAGCAGAGCTCA 1560
Qy 871 ValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGlyPhe 890
Db 1561 ATGGAATTCATATGACGAACTCGCGGTGCTCTTGGCACCTGATCATGAGGCGCG 1620
Qy 891 LeuThrValHisGlyAspIleValLysArgVal 901
Db 1621 GTGGTTATCCATGAGACAACTGAGGAGAGTT 1653

RESULT 6

US-10-128-714-7558
; Sequence 7558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7558
; LENGTH: 3072
; TYPE: DNA

! ORGANISM: Aspergillus fumigatus
US-10-128-714-7558

Alignment Scores:

Pred. No.:	2,23e-151	Length:	3072
Score:	1893.00	Matches:	438
Percent Similarity:	57.77%	Conservative:	157
Best Local Similarity:	42.52%	Mismatches:	278
Query Match:	40.85%	Indels:	159
DB:	5	Gaps:	23

US-10-768-511-8 (1-901) x US-10-128-714-7558 (1-3072)

```
Qy 12 ValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThr 31
   ::::|||||
Db 1 ATGCTCTCTCCAGCTCTCTCTCGCGTAGGGC----- 33

Qy 32 SerProSerPheAspAsnAlaAlaSerProValAlaGlyArgAlaValArgGlnThr 51
   ::::|||||
Db 34 CGCCAGCAAGGACACGCGCAACCTCTCTCCGCGCGGTCAACCCGTTCTCGTCAGCTC 93

Qy 52 ProThrSerAlaValArgArgGlyArgGlnThrAspSerAlaArgArgArgSer 71
   ::::|||||
Db 94 CAGACCAAGTA--GTCCGACTCCAGGGCGCGCATGAACAGTCAACGGCAGCCT--CGT 150

Qy 72 ArgSerArgSerLeuGlyAsnSerVal-----Tyr 81
   ::::|||||
Db 151 GCTTCAAGCGACTGAGAGCGGCAAGCGCGAGTTCCTCGTCATCGCCTATGTTCTCCAG 210

Qy 82 SerSerProTyrAspAlaGlyThrProGly---ThrProGlyThrProValAlaThrPro 100
   ::::|||||
Db 211 TCCTCCCTTCAAGGACAGACAGCAGTGGGAAACACACAGATGTTTCGAATGATGAGCCC 270

Qy 101 ValTyrAlaThrProVal-----GlyThrProMet 110
   ::::|||||
Db 271 -----AGTTCACCAATCGGAATCTTCTACTATGATGATGAAGGAGACAGGACCTAGA 324

Qy 111 Gly---ThrProSerPheHisArgGlyThrPro----- 120
   ::::|||||
Db 325 GGAATGCTCTTACCATGAGAGACTCTCTCCCATCCGGTATATGCTAGCTCCAGTCCA 384

Qy 121 -----GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeu 135
   ::::|||||
Db 385 ACTCGAGCTCAGAAATCGCGAGTCCCGCGGTTCCGACATTCACGAGCAGCAGCGGATTA 444

Qy 136 His-----ArgArgArgSerGlnSer 143
   ::::|||||
Db 445 TTCGTCTCGTCAGACCCAGCATCGAGACAACCGTGGCGGTGTCGCGCGTAGCGACCTT 504

Qy 144 ArgGluProGlyHis---ArgSerProSerArgGlu----- 154
   ::::|||||
Db 505 CATTCGTGGTGGTTTCTTCTAGCCGCAACCGTCCGCGGAGAGTGTTCGATGCTAAT 564

Qy 155 -----ProSerAlaAspGlyArgProSerGluSerAla-----GluPro 167
   ::::|||||
Db 565 GGTATGCTCGCAGCGATGCGATCCACGTTCCGATGCCACCTCTTCGAATATCCACCCA 624

Qy 168 Asp-----AspThrLeuGlyGlyGluTyrAla---TyrValTrpGlyThrAsn 182
   ::::|||||
Db 625 GACACCTCTGAGCGCGCGCTGGCGGTAGTCAACCCGTTGATTTGGGGTATCCAC 684

Qy 183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsn----- 198
   ::::|||||
Db 685 ATCTCCATCCAGATTCCTGCGCATTCAGAACTTCTCTACAACTTCCAAACAAA 744

Qy 199 TyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleLeuGluThrVal 218
   ::::|||||
Db 745 TACCGTCTGTGGCAGAGCGGCGCAACCGAGGATGAGACAGTATATGGGTGATTCAGCG 804

Qy 219 GluArgGluGlu-----AspThrLeu 225
   ::::|||||
Db 805 GAGGAACGGGAATATACATCAGCATGTTGAGCACCATCGCGCAACTTGGAGTAACCACTTG 864
```

```
Qy 226 AsnIleAspMetSerAspIle-----TyrAspHisAspProAspLeuTyrAlaLysIle 243
   ::::|||||
Db 865 AACTTGGATCGAAGAAGCTGAGGCATACCATCAACACTCAAGCTCAAGCATCAGTTG 924

Qy 244 ValArgTyrProLeuAspIleLeuProLeuLeuAspThrGluCysGlnGluValAlaThr 263
   ::::|||||
Db 925 CATGCTTATCTCAAGAGATCATTTCCATTGATGATCAGACAGTGAAGGATGTGATG 984

Qy 264 SerLeu----- 265
   ::::|||||
Db 985 GAGCTTGGCATCAAGGAGATGGAGCGCTCGGGCTCAGAACCAACCAAGAAATCAT 1044

Qy 265 ----- 265
   ::::|||||
Db 1045 AACAGAGCCTGAGCTCTGGTCCGCTGTTCCAGCTCTGACGGCTGAGTGAAACTGGC 1104

Qy 266 -----LeuProThrPheGluLysHisIleGluAlaArg----- 276
   ::::|||||
Db 1105 AGAATGCCACAGAACGAGATCCCGACCTTCTCGGTGAAGTGGAAACCAAGCATTCAA 1164

Qy 277 -----ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294
   ::::|||||
Db 1165 GTTCTGCTTTTGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCTCTGCAGCATG 1224

Qy 295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
   ::::|||||
Db 1225 GATAAATAGTAAGCATTAAGGGCTTAGTCTTGAACGACCCATCATTTCTCTGATATG 1284

Qy 315 LysGlyAlaPhePheLeuValCysGlyHisSerProProLeuValThrValVal 334
   ::::|||||
Db 1285 AAGAAGCTTCTTCCGTTGCAAGTCTGCAACCATGTTGTTTCAG--GTGACATTGAT 1341

Qy 335 LysGlyArgValGluProThrArgCysGluLysProGluCysAlaAlaAlaArgAsnAla 354
   ::::|||||
Db 1342 CGTGAAGATTCGGAACCCACTGAGTCCCGACCTCCAGTGTGTAAAGGAACAACTCG 1401

Qy 355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374
   ::::|||||
Db 1402 ATGCAACTCATCCATAACCGCTGTGTATTTCGCGCAAGACGAGTTCATCAAGTTGAGGAA 1461

Qy 375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394
   ::::|||||
Db 1462 ACACCTGACGATTCCTGATGGCGCAGACTCTCTACTCGGTTTCCTTGTGTGATGAT 1521

Qy 395 ThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAla 414
   ::::|||||
Db 1522 GAGCTGGTGGATGTCTGCAAGCTGTGTATCGGGTCAAGTCCAGCGTATTTCCGGTGC 1581

Qy 415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434
   ::::|||||
Db 1582 AACCTGTGCGCGTTAATCTCTCGCCAGCGGTACACAGAAAGTCTGTTCAAGACGTACATA 1641

Qy 435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
   ::::|||||
Db 1642 GATGTTCTTCATGTTTCAGAAAGATCGATCGAAGAAAGTGGGTATCGACGCTTCGACCATC 1701

Qy 455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla 474
   ::::|||||
Db 1702 GAGCAGGAGCTCTCGAACAGCGCTGGGGATGTCAGAACAAACACGTAGGCTCACTGG 1761

Qy 475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyLysTyrAspArg 494
   ::::|||||
Db 1762 GAGGAGGAA-----GAGAAATTAAGCGAACTGTCTACCAACCTGATCTGTATGAGCTT 1815

Qy 495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514
   ::::|||||
Db 1816 CTCTCTCGTCTCTGGCCCCCAGCATCTACAGATGAGACGCTGAAGAGGAATCTCTG 1875

Qy 515 CysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer-----PheArg 532
   ::::|||||
Db 1876 CTTTCTGTTGTTGGAGGACCAACACAGACCTTCCAGAGGGTGGTAACCCAGATACCTG 1935

Qy 533 GlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuGln 552
   ::::|||||
```


Qy	212	lnllelleGluGluThrValGluArgGlu	-----222
Db	419	GTAATAATGGGGGATTTCACGGCGAGGAACGGGAATACATCAGCATGTTGAGCACCATGCGGC	478
Qy	223	-----AspThrLeuAsnIleAspMetSerAspIle	-----TyrAspHisAspP
Db	479	AAC TTGAGTAAACACGAGTTGAACTTGGATGCGAAGAACCTGGAAGGCATACCCATCAACAC	538
Qy	237	roAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrG	257
Db	539	TCAAGCTATGGCATCAGTTGCATGCTTATCTCTCAAGAGATCATCTCCATTCATGATGATCAGA	598
Qy	257	luCysGlnGluValAlaThrSerLeu	-----265
Db	599	CAGTGAAGGATGTGATGGTGGAGCTTGGCATCAAGGAGATGGAGCGCTCGGGGCTCAGA	658
Qy	266	-----LeuProThrPheGlu	-----LysHisIleGluA
Db	659	ACCAACGAAACCAAGATCATAAACAGAGCGCTGAGCTCTGTGGAAACAACAGGCATTAACAG	718
Qy	275	laArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleA	295
Db	719	TTTCGCCCTTTGGTCTCACTCGACTGTGAATATGAGAGACCTCGATCTTCGACGACATGG	778
Qy	295	spLysLeuValSerValYsGlyMetValIleArgCysSerSerIleIleProGluIleL	315
Db	779	ATAAATAGTAAGCATTAAGGCGTTAGTTCATTCGAACGACACCCATCATCTCTGATATGA	838
Qy	315	ysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValL	335
Db	839	AAGAAGCTTTCTCCGTTGGCAACTGCAACCACTGGTGTTCAG- -GTGCAATGTGATC	895
Qy	335	ysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaM	355
Db	896	GTGGAAGATTGCGAACCCACTGAGTGCCACGCTCCAGTGTGTAAGCAAGCAAACTCGA	955
Qy	355	etSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluT	375
Db	956	TGCAACTCATCCATAACCGCTGTGTATTTCGCGAAGCAGGTCATCAAGTTGCAGGAAA	1015
Qy	375	hrProAspAlaIleProGluGluThrProHisIleThrValSerMetCysLeuTyrAsnT	395
Db	1016	CACCTGACAGCATTCCTGATGGCCAGACTCTCTCACTCGGTTTCCTTTGTGTATGATG	1075
Qy	395	hrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaM	415
Db	1076	AGCTGGTGGATGCTCGAAGGCTGGTGAATCGGGTGAAGTACCGGTATTTTCGGGTGCA	1135
Qy	415	etAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleA	435
Db	1136	ACCCTGTGGCGTTAATCTCGCCAGCGGTACACAGAAGTCGCTGTTCAAGACGTACATAG	1195
Qy	435	spCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluM	455
Db	1196	ATGTTCTTCATGTTTCAGAAGATCGATCGCAAGAAAGTTGGGTATCGACGCTCGACCATCG	1255
Qy	455	etAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaA	475
Db	1256	AGCAGGAGCTCTCGGAACAGCGCGCTGGGGATGCGAAGCAACAAACGTTAGGCTCTAC	1315
Qy	475	snGluAlaLysIleGlnLysLeuLysGlnLeuSerLysLeuProGlyIleTyrAspArgL	495
Db	1316	AGGAGGAA - - - - -GAGAAATTAAGCAACTGCTACCAAGACCTGTATGTATGAGCTTC	1369
Qy	495	euSerArgSerLeuAlaProSerIletrpGluLeuGluAspIleLysLysGlyLeuLeuC	515
Db	1370	TCTCTCGTCTCTGGCCCCCAGCATCTACGAGATGGACGACGTGCAAGAGGGAAATCTCGC	1429
Qy	515	ysGlnLeuPheGlyGlyLysAlaLysIleProSerGlyAlaSer - - - - -PheArgG	533
Db	1430	TTTCAGTTCTTTGGAGGCCACCAACAGACCTTCCAGAGGGTGGTAAACCCACGATACCGGTG	1489

Qy	533	lyAspIleAenValLeuValGlyAspProGlyThrSerIysSerGlnLeuLeuGlnT	553
Db	1490	GAGATATCAATATCTCTCTGTGGTGACCATCTCATCTCAAGTCCCAGTCTTTCGTT	1549
Qy	553	YrValHisValIleAlaProArgGlyIleYrThrSerGlyArgGlySerSerAlaValG	573
Db	1550	ACGTCCATAAGATTGCCCTTCGGGGTGATATACAGCGGCAAGGGCTCTCTCGCTGTG	1609
Qy	573	lyLeuThrAlaYrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA	593
Db	1610	GTCTTACGGCGTAGCTCACC CGCGCATCTCTGAACCCGCCAGATGGTCTCTCGAGTCGGGTG	1669
Qy	593	laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA	613
Db	1670	CCTTGGTTCCTTTCAGACGGCGGGTATCTGTGTGCATCGACAGTTCGACAAATCAACGAAT	1729
Qy	613	snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG	633
Db	1730	CCACTCGGTTCGGTCTCGATGAAGTCATGGAACACACAGACAGTATCTATCGCCAGGCAG	1789
Qy	633	lyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS	653
Db	1790	GCATTATCACTACTTTGAACGCTAGGACACAGCATCTCTGGCTTCGGCCATCCGATCGGTA	1849
Qy	653	erArgYrAsnAlaArgLeuSerValIleAspIleGlnLeuProProThrIleuLeuS	673
Db	1850	GCAGGTACAATCCCAACTTTCGCCGTCTCTCAAAATATTTGACCTTTCGCCCTTACCTTGTCTCT	1909
Qy	673	erArgPheAspLeuIleYrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgL	693
Db	1910	CCCGATTTCGACTTGGTATACCTCGTCTCGACCGAGTGGATGAGCAGGAAGATCGTCGGC	1969
Qy	693	euAlaArgHisLeuValAlaLeuHisYrGluAsn-----TyrGluValSerLysGlnA	711
Db	1970	TCGCTAAGCATCTTGTCAATATGTACTCGTGAAGACAGACACCTTGAGCATGCTGCCGAGCAAG	2029
Qy	711	spAlaLeuAspLeuGlnThrLeuThrAlaYrIleThrYrAlaArgGlnHisValHisP	731
Db	2030	AAATCTTTCGGCATCGAATTCCTTACAGCTTATATCACTTACGCCAAGCAACAAAGTCCATC	2089
Qy	731	roThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyYrValGluMetArgGlnL	751
Db	2090	CAGTGTCTACACCGCGCGCGTAAAGCCCTTGTGGATGCTTACGTTAACTGCGTTAAGC	2149
Qy	751	ysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluS	771
Db	2150	TTGAGATGACATCCGGTCTTCTCACCGCGTATCACCGCTACCATCTCGTCAACTGGAGT	2209
Qy	771	erMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysV	791
Db	2210	CCATGATCCGACTGTCGGAAGCGCATCGCGGTATCGGCTATCGCCGAGGTCACTCGCG	2269
Qy	791	alAspAlaAlaGluAlaValArgLeuAspValAlaLeuGlnGlnSerAlaThrAspH	811
Db	2270	ATGATGTGGAGGAAGCGCGTCTGATCCGCTCCGCCATCAAGACAGCGCGGCACCTGACT	2329
Qy	811	isAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgI	831
Db	2330	CTCGACCGGTCTGATCGACATGAGCTTGTTCGAGGGGACATAGTGTCCACGAGAGAC	2389
Qy	831	leArgArgAlaAsnLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProG	851
Db	2390	GCAGCCGGGAACACTCAAGCGTGCTTGTCTGTGGTGGATGATCTGTGCAGCGGTG	2449
Qy	851	lySerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerServ	871
Db	2450	CGCGTGCAGCT-----CGCTGGCGCGAGGTCTTCAGGATCTTAAAGCGAAGAACAGCAGCA	2503
Qy	871	alAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheL	891
Db	2504	TTGAGGTGGATGGAGCCCGATTTTCGGGATCGGTTTCGAGCGCTGGAGGCTGAGGAGCGG	2563
Qy	891	euThrValHisGlyAspIleValLysArg	900

```

Db      2564 TGAGTGTGTCGGCAGGCGGTGCGCGCG 2592
          :::::||||| :::::|||||
RESULT 8
US-10-425-115-98897
; Sequence 98897, Application US/10425115
; Publication No, US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98897
; LENGTH: 2939
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRP4577_21707C.1
US-10-425-115-98897

Alignment Scores:
Pred. No.:      6,57e-149      Length:      2939
Score:          1854.00      Matches:      426
Percent Similarity: 62.34%      Conservative: 145
Best Local Similarity: 46.51%      Mismatches: 244
Query Match:      40.23%      Indels:      102
DB:               8          Gaps:       24

US-10-768-511-8 (1-901) x US-10-425-115-98897 (1-2939)

Qy      20 SerGluGlyValSerThrProLeuProGlnValThrSerProSerPheAsnAlaAla 39
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      113 TCTGACGGCGGCGAGCTCCCTCTCCGATCGTCGCC--TATGGTCGGCGCTCA 169
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      40 SerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArgArg 59
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      170 AGCCACATCGCG-----GTTACTAATCTCTCCCATCTCAGCCACAGCGCGTCC 220
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      60 GlyArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSer 79
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      221 GGC-----GGTCGCGCTCGCGGTGTTCC----- 244
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      80 ValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThr 99
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      245 -----GCTAGTCTTAT---GCTTCGTCCTCCCTCTCTCGGGG-- 280
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      100 ProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThr 119
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      281 -----TTGAGACCGCG-----CCGACCTTGGCGCGCGCAG 313
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      120 ProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArg 139
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      314 CCG-----TCCGAGCTGGGGCT--GGCGCCCTCGCGAGCGCGGTCCAG 355
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      140 ArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGly 159
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      356 AACTCGACTGGCGGTTCCCG-----CCGACGCCCTCCACTCCAAATGTCCACCGATGAC 409
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      160 ---ArgProSerGluSerAlaGluProAspAspThrLeuGlyGly----- 173
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      410 GTCCCTCCATCTCCGAGACTGGGACGACGACGACGCGCGCGCGGCGGTCCAGCC 469
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      174 GluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArg 193
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      470 ACCCGGGTCTTCTGTCGGGACCAACAATAGCGTGCAGGACGTCAACGCGCCATTCTC 529
          :::::|||||:::|||||:::|||||:::|||||:::
Qy      194 ArgPheLeuHisAsnTyrArgSerSerAlaHisAspLeu----- 206
          |||||:::|||||:::|||||:::|||||:::|||||:::

```

```

Db      530 CGGTTCCTTGGCCCACTTCCGG---GACCCGCGCGACGCTGCGCGGTGCGACCCCGGTCAATG 586
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      207 ---AsnSerLysTyrIleGlnIleGluThrValGluArgGlu---GluAspThr 224
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      587 GACGAGGGCAAGTACATCGCGCCATCCACCGCATCTCTGAGCTCGAGGGCGGGAGTCG 646
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      647 CTCGACGTGCGATGCGCAGCGTGTTCACCCAGCAGCCAGACCTCTACAGCAAGATGTT 706
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      245 ArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSer 264
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      707 CGCTATCCGTCGAGGTCTCGCCATCTTCACATCGTGTCTCATGAGCATCGTCGCGCGC 766
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      767 ATCGAGCGCTCTTCGAGAGCACATCCAGACAGGATCTACACCTCAAGTCGTCATT 826
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      285 HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetVal 304
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      827 TGCTTGAGGAATCTCAACCATCTCATATTGAGAAGATGGTATCCATCAAGGTATGATA 886
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      305 IleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCys 324
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      887 ATTAGATGACCTCGGTACATACCGAGCTCAAGGAGCTGTGTCCGCTCGCTGTTGT 946
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      325 GlyHisSerProLeuValThrValLysGlyArgValGluGluProThrArgCys 344
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      947 GGTTCCTACTCAGAGCCGTCATGTTGATAGAGGAGAGTAACGTAACACACATTTGT 1006
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      345 GluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPhe 364
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1007 CAGAAAGAAATGTAAGCCACAAATCTATGACCTTAGTGCACACAGATGCAGATTT 1066
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      365 AlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThr 384
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1067 TCAGAACAGCAGATCAATAAGTTTCAGAAACACACAGACGAGATACCAAGAGTGCAC 1126
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      385 ProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAsp 404
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1127 CCACATCAGTTAGTCTTGATGCATGATAAGCTTGTGTGATGCTGGAAAGCTTGAGAT 1186
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      405 ArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArg 424
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1187 AGGTTGAGATAACTGGAATATACAGACTATGATATTCGAATTTGGACCAACTCAAAAG 1246
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      425 ThrLeuArgAlaLeuTyrLysThrTyrIle-----AspCysValHis 438
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1247 ACATGAAGTCTATTATCAAGACATATATCCCGGATGCGGTCTCTGAAGTGTGCGAG 1306
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      439 ValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGlu 458
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1307 GTGCGAGATGTTGTGGAGCCCTTCAGGCTTCTTGAAGTTGCCATGCGAG---CAGTCTGCA 1363
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      459 AsnAspMetTyrAlaGlyTyrHisGluSerAsp-----ThrSerGluAlaAlaAsn 475
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1364 ACGGATCATGCAACTGGGACGATTGATATGATATTAATCATGACGGGATATCTCAAGT 1423
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      476 Glu-----AlaLysIleGlnLysLeuLysGluLeuSer 486
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1424 GAAAGGACGAGACGGAACAGCTTGTGTCGCAACCCGTAACCTGATCGTGGAGAAATG 1483
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      487 LysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeu 506
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1484 CAGTCTGGAGGG-----CCTTCAATGCATATGATT 1513
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      507 GluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyLysAlaLysLysIlePro 526
          |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1514 GAGTTGCTGTA-AGAGCTCAGTCCGAGCTGTAATGAGAGCCATCTGCACGAATTCCT 1572
          |||||:::|||||:::|||||:::|||||:::|||||:::
Qy      527 SerGlyAlaSerPheArgGlyAspIleAsnValLeuValGlyAspProGlyThrSer 546
          |||||:::|||||:::|||||:::|||||:::|||||:::

```

```
Db 1573 CGGGCGCTTGGCACTATCATGACAGAAAG-----TGTGCTTCTCATCTCATGCAAAAAA 1626
QY 547 LysSerGlnLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGly 566
Db 1627 GGTGAAGAGAGGTTGATGTTAACT-----GCCATCTGCGAGGATGTT 1668
QY 567 ArgGlySerSerAlaValGlyLeuThrAlaTyr-ValThrLysAspProGluThrArgGly 586
Db 1669 TGTGTGAGCTCC-----GTCACCGCTTCTTCAACAGTCAGGTTGGCA 1713
QY 586 uThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGly 606
Db 1714 AGATCTGATCGACTAGGCGCATTTGCTTTTCAGAGTCGTAAAGTAGTTGTGTGATGATGA 1773
QY 606 pheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnTh 626
Db 1774 GTTTCATAGATGTCATTAATGCCGAGCATGTTACACGAGGTGATGGAACAGCAGAC 1833
QY 626 rValSerValAlaLysGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAl 646
Db 1834 AGTATCCATTGCGAAGCTGGAATAATTCATCTTTAAACGCTAGGACATCTGTCTGGC 1893
QY 646 aCysAlaIenProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGly 666
Db 1894 ATGTGCCAATCTCTACTGATCATCGTTTACAATCCAAAGGCTCTCTGTAATTGCAACATCCA 1953
QY 666 nLeuProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAs 686
Db 1954 CTTAGCGCCACGCTACTTCAAGATTGCACCTGATTTATCTATCTTTCGACACAGCGGA 2013
QY 686 pGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGly 706
Db 2014 TGAGCAAACTAGTAGCGCTCGCAAGCATATTGTTGTTGCAATTTTGAGAAATCCAAA 2073
QY 706 uValSerLysGlnAspAlaLeuAspLeuGlnThrLeuAlaTyrIleThrTyrAlaArg 726
Db 2074 TTTAGAGGAGCTCGAGTCTTGACCTTGACAGACACTAGTTTCTTACATAAGCTATGCAAG 2133
QY 726 glnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrVa 746
Db 2134 GAAGTATATTTCAGCCACAGTTATCTGATGAAGCTGCAGAGAGTTAACTCGTGCGCTATGT 2193
QY 746 lGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrPr 766
Db 2194 CAGATGAGAAAAGAGGAGGAATAGCCCTGGAGCAGAAAGAGGTCTATACAGCAACCGC 2253
QY 766 oArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGly 786
Db 2254 TAGACAAATAGAGAGTTTGCATCCGCTCTCAGCGAAGCATTTAGCCGGAATCGGTTCTCTGA 2313
QY 786 uValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGly 806
Db 2314 AGTGGTCGAGGTGCGGGATGTTGTGTGAGGCGATTTCAGGCTTCTTGAAGTCGCCATGCA 2373
QY 806 nSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrGlyValSer 826
Db 2374 GTCTGCGAGCGATCATGCAACTGTTAGTATGATGATGATGATGATGATGATGATGATGAT 2433
QY 826 rAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAs 846
Db 2434 CGCAAGCGAAAGCAGACGGCGGAGAACTCTGTTGCCGCAACCCGTAACCTGATTGCGGA 2493
QY 846 pLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArg 866
Db 2494 GAAATATGACGCTTGGAGGCGCCCTCG---ATGCGCATGATGATGATGATGATGATGATGAT 2550
QY 866 qSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGly 886
Db 2551 GAAGCAGAGCTCAATGGAATTCATATGACACGACATCCCGCGTCTCTTGGCACCCCTGAT 2610
QY 886 nGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
Db 2611 GACTGAGGCGCGGTGTTATTCATGAGACACACGTCGTGAGGAGATT 2656
```

RESULT 9

```
US-10-260-238-580
; Sequence 580, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 580
; LENGTH: 1723
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (35)..(95)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-580
```

Alignment Scores:

Pred. No.:	5,1e-148	Length:	1723
Score:	1840.00	Matches:	403
Percent Similarity:	64.41%	Conservative:	82
Best Local Similarity:	53.52%	Mismatches:	74
Query Match:	39.92%	Indels:	196
DB:	7	Gaps:	12

US-10-768-511-8 (1-901) x US-10-260-238-580 (1-1723)

QY	135	LeuHisArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGlu	154
DB	3	GTCCACCGACGACGCC-GGTTCTCTCGAGCGCGGCGGACGAGCACCC-----	52
QY	155	ProSerAlaAspGlyArgProSerGlnSerAlaGluProAspAspThrLeuGlyGlyGlu	174
DB	53	CCGAG-ACCGACG	99
QY	175	TyrAlaTyrValTyrGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArg	194
DB	100	CCGCTGTTCTGCTGCGGCGCACCAACATCAGCTGCGAGCGTCAACCGCGCATCTCCCG	159
QY	195	PheLeuHisAsnTyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIle	214
DB	160	TTCTTCGCGCACTTCGGACCGCGCG-----	186
QY	215	GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp	234
DB	186	-----	186
QY	235	HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu	254
DB	187	-----	198
QY	255	AspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleGlu	274

Db 199 TCGACC----- 204
QY 275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer-AspIle 294
Db 205 -----CGGTCAATGACGAGGCGCAAGATAT 228
QY 294 eAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
Db 229 CGAGAAGATGGGTGCATCAAGGGGTATGATTAATTCGGTGCAG-TCGGTGAATCCAGAGCT 288
QY 314 eLysGlyAlaPhePheLysCysLeuValCysGly----HisSerProProLeuValThr-V 333
Db 289 CAAGAGAGCTGTGTTCCGCTGCTGGTTCGGGTCTACTCTGAGCTGTGAATGGTTGA 348
QY 333 alValLysGlyArgValGluGluProThrArgCysGluLysP-ogluCysAlaAlaArgA 353
Db 349 TC-----GAGA 354
QY 353 snAlaMetSerLeuIleHisAsnArgCys--ThrPheAlaAsnLysGlnIleValArgLe 372
Db 355 ACTCTATGACTCTGTGCATACCGATCAGAGGGTTTCAGATTAAGCAGATCATAAAGTT 414
QY 372 uGlnGluThrProAspAlaIleProGluGluThrProHisThrValSerMetCysLe 392
Db 415 GCAGGAAACACCATGATGATACCAAGAGTGGCACTCCACATCTGTGTTGCTTGAT 474
QY 392 uTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPh 412
Db 475 GCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGGGTTGAG---ACTGGGATATA 531
QY 412 eLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysTh 432
Db 532 CAGAGCCATGAGTATTTAGAGTTGGCCCAACTCAGAGGACAGTCAAGTCGATTTTCAGAC 591
QY 432 rTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspPr 452
Db 592 GTACATTGATTCCTTCATANAAG----- 616
QY 452 oMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGl 472
Db 616 ----- 616
QY 472 uAlaAlaAsnGluAlaLysIleGlnLysLeuLysGlnLeuSerLysLeuProGlyIleTy 492
Db 617 -----AAGGAGTTGTCAAAAGTTGCCAGATATATA 645
QY 492 rAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGl 512
Db 646 TGACAGATTAACTAGTCAATTTGCTTGTGTTGCTCCAAACATATGGGAGCTGGACGCTTAAAGGGG 705
QY 512 yLeuLeuCys----GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPh 531
Db 706 CCTCTTTGCCAGCAGCTTTTGGTGGAAATGCTTTGAGGCTTCCTCTGGAGCTAGTTT 765
QY 531 eArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLe 551
Db 766 CCGAGGCGCACATCAATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 825
QY 551 uGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAl 571
Db 826 CCAATACATGCAACAACTGCTCTCGTGGCATTTATACAAGTGGCAGAGGAGTTTACG 885
QY 571 aValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSe 591
Db 886 TGTGGCCCTTACTGCATAGTTTACCAAGGATCCTGAAACTGGTGAACCTGTTCTTGAGAG 945
QY 591 rGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSe 611
Db 946 TGGAGCGCTGTGTTTGGTGGTGAACAAAGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1005
QY 611 rAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLys 631

Db 1006 TGATAATGCTCGAAGCATGTTTACATGAGGTGATGGAACAACAGACAGCTGCTCTCCATTGCCAA 1065
QY 631 sGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSe 651
Db 1066 GGCTGGAAATTAATTCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCAAAATCCTAC 1125
QY 651 rGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLe 671
Db 1126 TGAATCAGTTATTAATCCAGGCTCTCTGATTTGACATATCCATCTCTCTCCAACT 1185
QY 671 uLeuSer--ArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
Db 1186 GCTTTCTAGAGTTTGACCTCATTTATCTGATATTGGACAAGGCGAGATGAGCAAACTGAT 1245
QY 691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
Db 1246 AGAGCGCTGGTAAAGCATATTGTTTCGTTGCAATTTTGAGAAT----- 1287
QY 711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
Db 1288 -----CCAACTACATAAGTTATGCAAGGAAGCATATACAA 1323
QY 731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
Db 1324 CCACAGTTATCTGATGAAGCTGCAGAGAATTGACTCGCGCTATGTTGAGATGAGGAA 1383
QY 751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
Db 1384 AGAGAAACACGCCCTGGTAGCAGAAAGAG---ATAACTGGCAGACAGCTCGACAAATTGAG 1440
QY 771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
Db 1441 AGCTTGATTGGCTCAGTGAAGACATGCGCCGAATGCGATTCTCTGAATTTGGTGAAGTA 1500
QY 791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db 1501 CAAGATGTTGTAGAGGCTTCAGGCTTCGCAAGTTGCCATGCGACAAATCGCAACTGAT 1560
QY 811 HisAlaThr-GlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluAr 830
Db 1561 CATGCCACTGGTCAATCGATATGATCTTATCATGACTTGAATATCTCGAGCGAAAG 1620
QY 830 gIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPr 850
Db 1621 GCAGAGGCGG----- 1630
QY 850 oGlySerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSe 870
Db 1631 -----GACCAGTTGCTGGAGAAATTAGAGACAGAGCTC 1665
QY 870 rValAspValSerLeuGlnAspIle 878
Db 1666 TATGGAAGTTTCATCTCGCATGATGTA 1690
RESULT 10
US-10-320-797-2332
; Sequence 2332, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Broehkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2332
; LENGTH: 2970
; TYPE: DNA


```
Db 1840 GCCAAGGAGGATGATCATCCACTCTCAATGCTGTACATCTCTCGCCGCGGAAAC 1899
Qy 650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
Db 1900 CCAATCAACTCTCGATATGACCTTAACCTCCCTATTCCTGGCCACATCGATCTCCCTCT 1959
Qy 670 ThrLeuLeuSerArgPheAspIleuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
Db 1960 ACATTGATCTCAAGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2019
Qy 690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
Db 2020 GACAGAAAGTGGCAAGACATTTGCTGGAGCTCTATTTGTCAGATGATAGAGATCAGCT 2079
Qy 710 GlnAsp---AlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHis 728
Db 2080 GCCGACAATATCATCTCTTTCCAGACCTTAAGCTCTACATTCATATGCGCGCTCCAAA 2139
Qy 729 ValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMet 748
Db 2140 ATCCACCCGCTCTCCGCAAGAGCTTCTGAAGCGCTTGTTCAGGCGCTATGTTGAGATG 2199
Qy 749 ArgGlnLysGlyAsnPheProGlySerSerLysValIleThrAlaThrProArgGln 768
Db 2200 CGCAAGGCGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2259
Qy 769 LeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValVal 788
Db 2260 TTGGAAGTATGATCAGATTGGCGAAGCGCATGCGAGGATGAGATTAAAGTGACAGGGTT 2319
Qy 789 GluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAla 808
Db 2320 GAGGAAGAGGACATTAGGAGGCTGTTAGTTGATCAAGAGTGCCTCGAGGAGAGTGCT 2379
Qy 809 ThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSer 828
Db 2380 ACCGATCCCTTACTGGTCAAAATTGATCTTGACCTCAACACCGGTCGCGCGCCAAACC 2439
Qy 829 GluArgIleArgArgAlaAsnLeuAlaAlaLeuArgGluIleAlaAspLysIle 848
Db 2440 ATGCGTCGCGCCGAGCGCATCTTAAACGTGAAGTTATCAAACTCGTCGTGAAAG-- 2496
Qy 849 SerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGln 868
Db 2497 -----GCCCGATCTCAGGGTATTCGCTGGCGAGCAGTCATTGATGAGCTCAACAAACA 2550
Qy 869 SerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGlu 888
Db 2551 TCCAGTGTCCGTGGATCATGCACAGTTCGCAGAAATTGTTAGGAGCTGGAAGAGAG 2610
Qy 889 GlyPheLeuThrValHisGlyAspIleValLysArgVal 901
Db 2611 AGCATAGTCAAGGTATGCTGGTGAAGGAAAGAGGATT 2649
```

RESULT 11

```
US-10-369-46198
; Sequence 46198, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
```

```
; SEQ ID NO 46198
; LENGTH: 2802
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-46198
Alignment Scores:
Pred. No.: 3.95e-147 Length: 2802
Score: 1833.00 Matches: 393
Percent Similarity: 59.36% Conservative: 178
Best Local Similarity: 40.85% Mismatches: 281
Query Match: 39.77% Indels: 110
DB: 6 Gaps: 14
```

US-10-768-511-8 (1-901) x US-10-369-493-46198 (1-2802)

```
Qy 2 GluAsnAsnAspAlaLeuAspIleGlyAlaValSerProTyrProSerGlnSerGlu 21
Db 31 GAGGATAATACTCC-----AGCTCCCGTTGTGCTTAATCTCTGAT 72
Qy 22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerPro 41
Db 73 TCTGTTCCTCA-----CCACAGCTTCTTCCCGAGCTCTATTTTATAGCTCTCTTCA 123
Qy 42 ValAlaGly----- 44
Db 124 TCACAAGGTGATATCTATGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGAGGAAAC 183
Qy 45 ---ArgArgAlaValArgGlnThrPro-----ThrSerAlaValArgArgGly 60
Db 184 ATCAGAGCTGATAGGTTCTTCTCCATAAATTTTCCATCTTCTCCCAAGACAAAAT 243
Qy 61 ArgGluThrAspSerAlaArgArgArgSerArgSerLeuGlyAsnSerVal 80
Db 244 TCCGATGTTTCCCAATCTCAAGGCAGACAGGCGAGAAATTCGTTCTCTCGCGCGCTTCT 303
Qy 81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db 304 GGAAGGTCTAGATATCATCTCTGATCTGAGAAGTGAAGACGCTGCTTCTTCTTCC--- 360
Qy 101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db 360 ----- 360
Qy 121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
Db 361 -----TCTTCTTTAGGCGGTAATGTCCTCAAAACCGGTGTACACATGCGAAGA 405
Qy 141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db 406 AATGATATTCATACATCTGATTTATCTCTCCAGAAGAATTGCGATTTTGATACTAGA 465
Qy 161 -----ProSerGluSerAlaGlu 166
Db 466 TCCGGGTGATACCTTTGGATACCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 525
Qy 167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValThrGlyThrAsnValAsnIlePro 186
Db 526 CCC-----TTGAGAATAAATTTGGGGTACCAACGCTCAGTATCCAG 564
Qy 187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg----- 200
Db 565 GAATGTACAACTAATTTTCGTAATTTTGTGATGCTTTTCAAGTATAAATTCGTTAAATA 624
Qy 201 -----SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
Db 625 TTGGATGAAGGAGGAATTCATTAAACAATACCTACCGACGAGAACTATATCTATCAAG 684
Qy 213 IleIleGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
Db 685 CAGCTTAATGAATGAGAGAAGCTTGGTACCTCTAATTTTAAACTTGGATGCTAGAACTTA 744
Qy 233 -----TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
```


; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 27488
 ; LENGTH: 2881
 ; TYPE: DNA
 ; ORGANISM: Neurospora crassa
 US-10-369-493-27488

Alignment Scores:
 Pred. No.: 6,57e-146 Length: 2891
 Score: 1819.00 Matches: 419
 Percent Similarity: 58.17% Conservative: 161
 Best Local Similarity: 42.03% Mismatches: 249
 Query Match: 39.47% Indels: 169
 DB: 6 Gaps: 24

US-10-768-511-8 (1-901) x US-10-369-493-27488 (1-2881)

Qy	24	SerThrProLeuProGlnValThrSerProSerPheAsp-----	36
Db	1	AGCTCGCTTTGGCCGACAGACAGACAGACCGATGTCACCGAAGTCTCTCGCGCA	60
Qy	37	AsnAlaAlaSerProValAlaGlyArgAlaValArgGlnThr-----	51
Db	61	AACGACGAGGACAGTAAATTATTTGGGGGTATGGTGTCAACCAAAATAAAAAAAGGAC	120
Qy	52	-----ProThrSerAlaValArgArgGlyArgGluThrAspSerAla	66
Db	121	TACACTGAATGCTGACCACTCCGCA-----	156
Qy	67	ArgArgArgSerArgSerArgSerLeuGlyAsn-SerVal.TyrSerSerProTyr--	85
Db	157	CCCATCCGCTATGCTCCAGTTCTTAGCCGGCCGTCAGCTTACTCAACAGTCCGATCTT	216
Qy	86	-----AspAlaGlyThrProGlyThrProGlyThrProValAlaThrProVa	101
Db	217	CGTAGCGAAAGCAGTCAGCTCTTCGTACGTCGCAAGGTCCGTCGCGTCGATCCGCT	276
Qy	101	lTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGl	121
Db	277	CGAGCGACATCAACGGGAC-CCTCTTCGTACTCCCGCC-----CA	317
Qy	121	nTyrIysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSe	141
Db	318	GATTCGCCCGAATCATCTTCGACGACGCGCGTGTGTATCGC-----	363
Qy	141	rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr	161
Db	364	-----GATGCCCTCGCTCCGATGCCAATC-----	389
Qy	161	oSerGluSerAlaGluProAsp-----AspThrLeuGlyGlyGlu---TyrAl	176
Db	390	GTTCGTACCAATAACCCCAATACCTCTGAGCGAGATGCCCTTGGTGACAAAGCCAAAG	449
Qy	176	aTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgPheLe	196
Db	450	TCTTGTGGGTACAAATTTCTTTCGACGATTCTCTCCGGTTCGATCAAGGACTTCT	509
Qy	196	uHisAsn-----TyrArgSerSerAlaHis-----	204
Db	510	CAGGAATCTCACAGGAATACCGGATGTGGCAGATGGCGGCGAGCGCGAGACTAT	569
Qy	205	-----AspLeuAsnSerLys---TyrIleGlnIleLeuGluThrValGluAr	220

Db	570	TGGCCATCCGGATCGGATTCGAAGCCATATTTGGGAGGCACTGGAGAATATGCTGCTGCT	629
Qy	220	gGluGluAspThrLeuAsnIleAspMetSerAspIle-----TyrAspHisAspProAs	238
Db	630	TGGTACCAACAAGCTGTATCTGGATCTCCGCGACTCAAGTCTTACCCGCGCACCTTGA	689
Qy	238	pLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThr----	256
Db	690	GTTATGGATCAGGCCCAACACTATCTTACCGAATCATCCCGTCAATGGATCAATGCGT	749
Qy	257	-----GluCys-----GlnGluValAlaThr-----	263
Db	750	ACATGACTGTATGATGGGCGCAAAAGGAATGGCCAGCCCAAGAGCATCCCAAA	809
Qy	263	-----	263
Db	810	CTCTAGGACCGTCCAGGTGCTCACAAGCTCGGAGCCGAACTTCCCAAGCTCTGAAG	869
Qy	264	-----SerLeuLeuProThrPheGluLysHi	272
Db	870	GAGCGAGAAACCGCCACCGCGCTGCCCAACAGCGCGCCCACAATTTGAGGACCA	929
Qy	272	sIleGlu-----AlaArgProPheAsnLeuLysAlaSerValHisMetAr	287
Db	930	AGTCAGCCAGATGGCGTATGTGGTTCGCCCGTGGGGTCTTGACAAGATCACCACCTCG	989
Qy	287	gGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCy	307
Db	990	TGACCTCAACCATCTGACATGGACAAGCTTGTGTCCATCAAGGTCTTGTTCATCAGA	1049
Qy	307	sSerSerIleIleProGluLysGlyAlaPhePheLysCysLeuValCysGlyHisSe	327
Db	1050	CACCGCGCTCATCCCGACATGAAGAGTCAATTTTCAAGTGCATGTGTTGGTGCATTC	1109
Qy	327	rProProLeuValThrValVal-----LysGlyArgValGluGluProThrArgCysGl	345
Db	1110	T-----ATAACCGTTCAGTTAGATCGTGGCAAGATCTGTGAGCTTACAGAGTGTCC	1160
Qy	345	uLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAl	365
Db	1161	TCGGGCTCGCTCGCTCCAAAGAACTCAATGCAGATCATCCACAACAGATGTGCTTTCGA	1220
Qy	365	aAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPr	385
Db	1221	AGACAAGCAAGTCAATCAAGCTGCAAGAACTCCCGATAACCTGCTGTGTCGACACC	1280
Qy	385	oHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspAr	405
Db	1281	CCATCGGTATCCGTTGGCTTATTAACGACTTGTGGACTTTTTCGAAGGCGGTGACAG	1340
Qy	405	gIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgTh	425
Db	1341	AGTTGAGTCACTGGTATCTTTAAAGTCACCCCGTTCGTCAACCCGCGCATGAGGAC	1400
Qy	425	rLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGl	445
Db	1401	GGTCAAGAGCGTCGACAAGACATATGTGTATGTTGTCTCCATACAAAAGGTCGACGAA	1460
Qy	445	YArgLeuGlnThrGluAspProMetGluMetAsp-----LysGluAsnAspMetTyrAl	463
Db	1461	GAGATGGTGTCT---GATCCCTCGACCTCGACTTCCCGAGGAGGAGGCCCATGC	1517
Qy	463	aGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLy	483
Db	1518	TAATGGCCAGAGCATGGCAAGTCCGAAAGGTCTCTCTGATGAGGAAGAGAGATCAA	1577
Qy	483	sGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIl	503
Db	1578	GGAACTCCCGCTCGACCGGATATCTACGATCTCTCTTTCACGTCTTCTTGGGTCCATCA	1637
Qy	503	eTrpGluLeuGluAspIleLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLy	523
Db	1638	TTACGAGATGATGTGAAGAGGTATTCTTCTTCACTGTTCGAGGCGCACCAACAA	1697

```
QY 523 slyslleProSerGlyAlaSer-----PheArgGlyAspIleAsnValLeuValG1 541
Db 1698 GACCTTTGAGAAGGCTGGAGGCCCAATAACCGTGAGACATCAAGTTCCTCTGTGG 1757
QY 541 yAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgG1 561
Db 1758 TGACCATCCCACTTCCAAAGTCCAGCTTCTCTCTATGTTCACAAATCGCCCAAGAG 1817
QY 561 yIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAs 581
Db 1818 TGTGTACAGAGTGTGAAGGTCTCTCCCGTAGGTCTTACCGCTATGTACCGCA 1877
QY 581 pProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly1 601
Db 1878 TCCCGAGTCCCGCAACTCGTTCTAGAGTCGGTGTCTCTCGTCTTATCGACGGCGTGT 1937
QY 601 eCysLysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluVa 621
Db 1938 CTGTTGTATCGACGAGTTCGACAAATGAACGAATCGACCCGCTCCGTTCTTTCACGAAGT 1997
QY 621 lMetGluGlnGlnThrValSerValAlaLysGlyIleIleAlaSerLeuAsnAlaAr 641
Db 1998 CATGAACAACAGACGCTGTCGCGCAAGCGCGCATCATCACCACTCTCAATGCCCG 2057
QY 641 gThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerVa 661
Db 2058 AACATCATCTCGGCTGCGCAACCCCAATGGTAGATACAAACCTGATCATCAT 2117
QY 661 lIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrIleuMe 681
Db 2118 GCGCAAAATATGACCTCCCGCAACCTTCTATCCCGTTTCGATCTCGTCTACCTCAT 2177
QY 681 tLeuAspLysProAspGluGlnAsnAspArgLeuAlaArgHisLeuValAlaLeuHi 701
Db 2178 CTTGATCGTGTGTGAAGAAATGATCAGCGTCTGGCCCGCATTTCTCTCTCATGTA 2237
QY 701 sTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu----- 715
Db 2238 TCTCGAAGACAGCCGAGGTGCCGAGCAGGCAACGAGTCTTGTGAGTTTTTCCCC 2297
QY 715 ----- 715
Db 2298 GTCTGTCAATCCACAGAGATGAACCCCAACTAACCACTCCCCCTCAAAAAAACAGC 2357
QY 716 ----GlnThr-LeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisPProThrLeuS 734
Db 2358 CCGTCGAGTTTAACTCTCTACATCTCTACGCGCGCTCGCACATCCACCGGCGCTCA 2417
QY 734 eAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnP 754
Db 2418 CGCCCAAGCGGTCCGAGCTCGTCGATCGCTACGTCGATGCGCAAGCTGGGCAAG 2477
QY 754 heProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleA 774
Db 2478 ACGTGGCGCGCGAGAGCGCATTAACGCCACCAACAGTCAGTCTCGATCCATCATTC 2537
QY 774 rGleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaA 794
Db 2538 GCTTAGCCGAGGCGCAGCAAGATGGTTTATCTCAAAACCGTCACGCGCGACGCTCC 2597
QY 794 lagluAlaValArgLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrG 814
Db 2598 GCGAGCGCGTGCCTCATCAAGTCGCGCTCAAGACCGCGCCGACCGAC---AGCAGG 2654
QY 814 lYThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgA 834
Db 2655 GCGCATTCGACATGAGCTTGTTCACAGAGGCGACGAGCGCGGCGGAGCGCCAGCAAG 2714
QY 834 laAsnLeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSers 854
Db 2715 CCGACATGAAGATCGGTTTATCCGCTGTGTGATGAGATGACACGCGCGCGAGTGA- 2773
```

```
QY 854 erGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValS 874
Db 2774 -----GTTAGTACTCGAGGTGGCGAGACAGCGTTTGGCGGAGCGGCGGTGTGCAGGTTG 2828
QY 874 erLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlnGlyGluGly 889
Db 2829 AGCCGGCTGAGTTTCCGAGGTTATGAGGGCTCTGGAGATGGAGGCG 2875

RESULT 13
US-09-954-456-568
; Sequence 568, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 568
; LENGTH: 3273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-568

Alignment Scores:
Pred. No.: 1,05e-145 Length: 3273
Score: 1817.50 Matches: 406
Percent Similarity: 57.78% Conservative: 151
Best Local Similarity: 42.12% Mismatches: 256
Query Match: 39.43% Indels: 151
DB: Gaps: 18

US-10-768-511-8 (1-901) x US-09-954-456-568 (1-3273)

QY 38 AlaAlaSerProValAlaGlyValArgAlaValArgGlnThrProThrSerAlaValArg 57
Db 21 GCGACGTCGTCGCGAAGCGCGCGC-----CTTCCAGTAACGCG---CGC 62
QY 58 ArgArgGlyValArgGluThrAspSerAlaArgArgArgSerArg----- 72
Db 63 CGCGGGGGGAGGCGGTGGCGGCGAGCCGCGGAAGCTCGCGGTCGCGAGCAGGCGCAG 122
QY 73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db 123 GGAAGCGCGGAGCGCGC---CCGCGCCGAGCTTGTCTTGTGCGCGCATCTCCGAGC 179
QY 92 ThrProGlyThrProValAlaThrProVal-----TyrAlaThr 104
Db 180 ACTATGTCGTCGCCGCGCTCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACC 239
```

Qy	105	ProValGlyThrPro 	-----	109
Db	240	CCCGCCAGACGCCTCGAGTGGAGATGCCAGGTCACTCCCTCTCAGAGACGTAGAGGC	299	
Qy	109	-----	109	
Db	300	GAGGATTCCACTCCACGGGGAGTTGCAGCCGATGCCAACCTCGCTGGAGTGGACCTG	359	
Qy	109	-----	109	
Db	360	CAGAGCACTGTCGCGCAGGACGTCTGTTTTTCAGCCCTCCCCAATGCATCTTCTAGCT	419	
Qy	110	-----MetGlyThrProSerPheHis--	116	
Db	420	ATCCCTCTTGACTTTGATGTTAGTTACCACCTGACATACGGCATCTCCAGCTCTCGGGA	479	
Qy	117	-----ArgGlyThrProGlnTyrlsGlnArgSerGluLeu	128	
Db	480	GAGGGAAACCCAGAAGTGGTTAGGGGCACACT--GTGAGACAGAGCCCTGACCTG	536	
Qy	129	GlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluProGlyHis	148	
Db	537	GGCTCTGCACAGAGGGCCTGCAGTGCATCTGCAGTCT	575	
Qy	149	ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp	168	
Db	576	-----GACGGGGCACGACGACAAGATATAGTGGCAAGT	608	
Qy	169	AspThrLeuGlyGlyGluTyralTyrrvalTrpGlyThrAsnValAsnIleProAspVal	188	
Db	609	GAGCAGTCTTAGGCCAAAACCTTGTGATCTGGGGAACAGATATAATGTGGCAGCATGC	668	
Qy	189	LeuArgAlaIleArgArgPheLeuHisAsnTyr-----	200	
Db	669	AAAGAAACCTTCAGAGATTCTTCAGCGTTTATTGACCTCTGGCTTAAGAGAGAA	728	
Qy	201	SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu	219	
Db	729	AATGTTGGCATAGATATTACTGAACCTCTATACATGCACGATGGGGAGATTAAATGTT	788	
Qy	220	ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrIspHisAspProAspLeu	239	
Db	789	ATTGGTGAGCAATTTTAAATGTGAAGTGTGAACACATCAAATCATTTGACAAAAATTG	848	
Qy	240	TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln	259	
Db	849	TACAGACAACCTCATCTCTTCCCACAGAGAGTTATCCAACTTTTGACATGGCTGTCAAT	908	
Qy	260	GluValAlaThrSerLeuLeuPro-----ThrPheGluLysIleIleGluAlaArgPro	277	
Db	909	GAAATCTTTTGACCGTTACCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA	968	
Qy	278	PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu	297	
Db	969	TTCAACGCATTGAAGACCTAAGAAATATGAGAAACCTGAATCCAGAAGACATTTGACCACTC	1028	
Qy	298	ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla	317	
Db	1029	ATCACCATTACGGCGATGGTGATCAGGACATCCCGCTGATTCGGAGATGCGAGAGGCC	1088	
Qy	318	PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg	337	
Db	1089	TTCTTCAGTCCAAGTGTGTGCCACAG---ACCCGGGTGGNAGTGGACCCGGCCGC	1145	
Qy	338	ValGluGluProThrArgCysGlyLysProGluCysAlaIalaArgAsnAlaMetSerLeu	357	
Db	1146	ATTCAGAGCCCGAGTGTGTGGCGGC-----TGCCACACCACTCCACGATGGCACTC	1199	
Qy	358	IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp	377	
Db	1200	ATCCAAACCGCTCCCTCTCTCTGACAGAGCATGATCAAGCTTCAGGAGTCTCCGGAA	1259	
Qy	378	AlaIleProGluGlyGluThrProHisThrValSerMetCysLautyrAsnThrMetVal	397	

[illegible]

Db 2295 GAAGCCAGCCAGCTCTCATCGAGGCTTATGTAGATGAGGAAG-----ATT 2342

QY 756 GlySerSerLysValLeuThrAlaThrProArgGlnLeuGluSerMetLeuArgIle 775

Db 2343 GGCAGTAGCCGGGATGTTCTGCTATACCTTCGACAGCTAGATCATTAATCCGCTTA 2402

QY 776 SerGluAlaLeuAlaArgMetArgPheSerGluValGluLysValAspAlaGlu 795

Db 2403 GCAGAGCCCATGCTAAAGTAAGATTGCTCAACAAAGTTGAAGCCATGATGTGAAGAG 2462

QY 796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815

Db 2463 GCCAACCGCTCATCGGGAAGCTCTGAAGCAGTCTGAACATGATCCCGAGTGGCATC 2522

QY 816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgAlaAen 835

Db 2523 GTGGACATATCTATTCTTACTACGGGATGAGTGCACCTCTGTAACGGGAAGAGAA 2582

QY 836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSergly 855

Db 2583 TTAGCTGAAGCATTTGAAGAAGCTTATT-----TTATCTAAGGGCAAAACACACAGCT 2633

QY 856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875

Db 2634 CTAAATATCCGCAACTTTTGAAGATATTCGGGGCAATCTGACATAGCAATTAATAA 2693

QY 876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895

Db 2694 GATATGTTGAAGAGCACTGCTGCTGCGGCGATGATGATTTCTTGACATGACTGGG 2753

QY 896 AspIleValLys 899

Db 2754 AAGACCGTGGCG 2765

RESULT 14

US-09-954-456-569

; Sequence 569, Application US/09954456

; Patent No. US20020115057A1

; GENERAL INFORMATION:

; APPLICANT: Young, Paul

; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

; FILE OF INVENTION: Sats

; FILE REFERENCE: 689290-76

; CURRENT APPLICATION NUMBER: US/09/954,456

; CURRENT FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: US/60/233,617

; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: US/60/234,052

; PRIOR FILING DATE: 2000-09-20

; PRIOR APPLICATION NUMBER: US/60/234,923

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,134

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,637

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US/60/235,638

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US/60/235,711

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,720

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,840

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,863

; PRIOR FILING DATE: 2000-09-27

; NUMBER OF SEQ ID NOS: 2276

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 569

; LENGTH: 3273

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-954-456-569

Alignment Scores:

Pred. No.: 1.05e-145 Length: 3273

Score: 1817.50 Matches: 406

Percent Similarity: 57.78% Conservative: 151

Best Local Similarity: 42.12% Mismatches: 256

Query Match: 39.43% Indels: 151

Db: 3 Gaps: 18

US-10-768-511-8 (1-901) x US-09-954-456-569 (1-3273)

QY 38 AlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaValArg 57

Db 21 GCGAGCTGTGGCAAGCGCGCG-----CTTCCAGGTAAACGG---CGC 62

QY 58 ArgArgGlyArgGluThrAspSerAlaArgArgArgSerArg-----72

Db 63 CGCGCGGGGAGGGCGTTGCGCGGAGCGAGCGGGAACGTCCGCGTGCAGGAGCGGCGAG 122

QY 73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91

Db 123 GGAAGCCCGGAGCGCGGC---CCGCGCGGAGCTTGTCTTGTGCGCGAGGTACTCGAGC 179

QY 92 ThrProGlyThrProValAlaThrProVal-----TyrAlaThr 104

Db 180 ACTATGTCTGTCCTCCCGCGTGCACCCCGAGCCCGCGGCGCGCGCGTGAAGGCCACCC 239

QY 105 ProValGlyThrPro-----109

Db 240 CCGCGCCGAGACGCTCGGAGTGAGGATGCCAGGTATCTCCCTCTCAGAGACGTAGAGGC 299

QY 109 -----109

Db 300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCGGATGCCAACCTCGCTGGAGTGGACCTG 359

QY 109 -----109

Db 360 CAGAGCACTGCTGGCGAGGACGTGCTGTTTTCAGCCCTCCCAATGATCTTTCAGCT 419

QY 110 -----MetGlyThrProSerPheHis---116

Db 420 ATCCCTCTTGACTTTGATGTTAGTTTACCCACTGACATACGCGCATCCCGCTCTCGGTA 479

QY 117 -----ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128

Db 480 GAGGGAAACCCCAAGTGGTGTAGGGGCACACCT---GTGAGACAGAGCGCTGACCTG 536

QY 129 GlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluProGlyHis 148

Db 537 GGCCTCTGCACAGAGGGCCCTGCAAGTGTGATCTGCAGTCT-----575

QY 149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168

Db 576 -----GACGGGGCAGCAGCAGAGATATAGTGGCAAGT 608

QY 169 AspThrLeuGlyGlyLeuTyrAlaTyrValTyrGlyThrAsnValAsnIleProAspVal 188

Db 609 GAGCAGTCTCTAGGCCCAAAAACCTTGATCTGGGGAACAGATGTAATATGTGCGAGCATGC 668

QY 189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----Arg 200

Db 669 AAAGAAAACCTTTCAGAGATTTCTTCAGCGCTTTTATTTGACCTCTGCTGCTAAAGAGAA 728

QY 201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleGluGluThrValGlu 219

Db 729 AATGTTGGCATAGATATTAATCTGAACCTCTATATCATCAGCATCTGGGAGATTAATGTT 788

QY 220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239

Db 789 ATTGTTGAGCAATTTTAAATGTGAACCTGTGAACACATCAATCAATTTTGACAAAAATTG 848

QY 240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259

Db 849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTTCCAACTTTTGACATGGCTGCAAT 908


```

, PRIOR APPLICATION NUMBER: US/60/234,052
, PRIOR FILING DATE: 2000-09-20
, PRIOR APPLICATION NUMBER: US/60/234,923
, PRIOR FILING DATE: 2000-09-25
, PRIOR APPLICATION NUMBER: US/60/235,134
, PRIOR FILING DATE: 2000-09-25
, PRIOR APPLICATION NUMBER: US/60/235,637
, PRIOR FILING DATE: 2000-09-26
, PRIOR APPLICATION NUMBER: US/60/235,638
, PRIOR FILING DATE: 2000-09-26
, PRIOR APPLICATION NUMBER: US/60/235,711
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: US/60/235,720
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: US/60/235,840
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: US/60/235,863
, PRIOR FILING DATE: 2000-09-27
, NUMBER OF SEQ ID NOS: 2276
, SOFTWARE: PatentIn version 3.0
, SEQ ID NO 976
, LENGTH: 3273
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-954-456-976

```

Db					:::	-----	578
Qy	537	GGCTCTGCACAGAGGCGCTGCAGGTGATCTCGAGTCT					
Db					:::	-----	579
Qy	149	ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp					
Db	576	-----	GACGGCGCAGCAGCAGAAGATATAGTGCAAGT				
Qy	169	AspThrLeuGlyGlyGluTyraIaTyValTTPGLYThrAsnValaenlleProAspVal					
Db	609	GACAGTCTCTAGCCCAAACAACCTTTGTGATCTGGGAACAGATTAAATGTGGCGCATGC					
Qy	189	LeuArgAlaIleAargPheLeuHisAsnTyr				-----Arg	200
Db	669	AAAGAAAACITTCAGAGATTCTTCAGCGTTTTATTGACCCCTCTGGCTAAAGAAGAAAGAA					728
Qy	201	SerSerAlaHisAspLeuasnSerLys--TyrileGlnlleleleGluGluThrValGlu					219
Db	729	AATGTTGGCATATATTAAGAACCTCTATACATGCAACGACTTGTGGGAGATTAATGTT					788
Qy	220	ArgGluGluaspThrLeuasnilleAaspMetSerAspileTyrAspHisAaspProAspLeu					233
Db	789	ATTGGTGAGCAATTTTAAATGTGAACATGTGAACACATCAATCATTTTGACAAAAAATTTG					848
Qy	240	TyrAlalysileValArgTyrProLeuAspileleProLeuLeuAspThrGluCysGln					259
Db	849	TACAGACAATCATCTTACCACAGGAAGTTATCCAATTTTGACATGGCTGTCAAT					908
Qy	260	GluValalaThrSerLeuLeuPro-----ThrPheGluIlysHsilleGluUlaAargPro					277
Db	909	GAATCTCTTTGACCGTTACCTCGACTCAATCTTAGAACATCAGATTCAAGTAGAACCA					968
Qy	278	PheAsnLeuIysAlaserValHisMetArgGluLeuasnProSerAspileAspLysLeu					297
Db	969	TTCAACGCATTGAAGCATTAAGATATAGAAAACCTGAATCAGAAGACATTCAGCACGTC					1028
Qy	298	ValserVallysGlyMetValIleAargCysSerSerielleProGluIleLysGlyAla					317
Db	1029	ATCACCATCAGCGCATGTGTATCAGACATCCCAGCTGATTCGAGATGACGAGGCC					1088
Qy	318	PhePhelysCysLeuValCysGlyHisSerProProLeuValThrValLysGlyAarg					337
Db	1089	TTCTTCCAGTCCAAAGTGTGTGCCACACG---ACCCGGTGGAGATGAGCCGCGCGCGC					1145
Qy	338	ValGluGluProThrArgCysGluLysProGluCysAlaIalaArgAsnAlaMetSerLeu					357
Db	1146	ATTGCAGAGCCCATGTGTGTGGCGCGC-----TGCCACACACCCACGATCGCATC					1199
Qy	358	IleHisasnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp					377
Db	1200	ATCCAAACCGCTCCCTCTTCTCTGACAGCAGATGATCAAGCTTCAGAGTCTCCGGA					1255
Qy	378	AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal					397
Db	1260	GACATGCTCGAGGCGCAGACACACACACAGTTATCTCTGTGTGTCACATGATCTCGT					1315
Qy	398	AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal					417
Db	1320	GACAAGTCCAGCTCGGGGACAGAGTGAATGTACAGGCATCTATCCAGCTGTGCTTAT					1379
Qy	418	ArgValGlyProasnGlnaArgThrLeuArgAlaLeuTyrIysThrTyrIleAspCysVal					437
Db	1380	CGAGTCAATCCACAGATGAGTAATGTGAAGTCTGTCTACAAAACCCATTTGATGTCAAT					1439
Qy	438	HisValLysIysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys					457
Db	1440	CATTATCGAANAACGATGCAAAACGTTGCAT-----					1472
Qy	458	GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla					477
Db	1473	-----GGCCTTGATGAAGAAGCAGAAACAGAACTTTTTCAGAGAA					1514
Qy	478	LysIleGlnIysLeuLysGluLeuSerIysLeuProGlyIleTyrAspArgLeuSerArg					497

```
Db 1515 CGTGTGGAAATGCTTAAGGAACTTTCCAGAAACACAGACATTTATGAGAGGCTTCCTTCA 1574
Qy 498 SerLeuAlaProSerIleTyrGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
   ::::::::::::::::::::|
Db 1575 GCCTGGCTCCAAGCAATTATGAACATGAAGATAAAGAG3GAATTTTGGCTTCAGCTC 1634
Qy 518 PheGlyGlyLysAlaLysLysIle-----ProSerGlyAlaSerPheArgGlyAspIle 535
   ::::::::::::::::::::|
Db 1635 TTTGCGGGCAACAAGGAAGATTTTAGTCACACTGGGAAGGGGCAAAATTTCCGGGCTGAGATC 1694
Qy 536 AsnValLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
   ::::::::::::::::::::|
Db 1695 AACATCTTGTGTGTGGGACCCCTGTGACCAGCAAGTCCAGCTGCTCGCATGACGTGTAC 1754
Qy 556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
   ::::::::::::::::::::|
Db 1755 AACCTCGTCCCGAGGGCCAGTACACGCTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814
Qy 576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
   ::::::::::::::::::::|
Db 1815 GCGTACGTAATGAAGAAGCCCTGAGACAAGGCGAGCTGGTCTCTGACAGAGGTGCTCTGTGC 1874
Qy 596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysIleMetSerAspAsnAlaArg 615
   ::::::::::::::::::::|
Db 1875 CTGAGTGAACACGGCATCTGCTGTATCGATGATTCGACAGTGTGACAGTGAAGTACAAGA 1934
Qy 616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyIleIle 635
   ::::::::::::::::::::|
Db 1935 TCGGTATTGTCATGAAGTCATGGAACAGCAGACTCTGTCTCATTTGCAAGAGCTGGGATCATC 1994
Qy 636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
   ::::::::::::::::::::|
Db 1995 TGTCAGCTCAATTCGCGCAGCACTCTGCTCTGGCAGCAGCAAAATCCCATTTGAGTCTCAGTGG 2054
Qy 656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThrLeuLeuSerArgPhe 675
   ::::::::::::::::::::|
Db 2055 AATCCTAATAAACAACCAATTGAAACATCCAGCTGCCTCATCTTTATTATCAAGGTTT 2114
Qy 676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
   ::::::::::::::::::::|
Db 2115 GATTTGATCTTCCTCATGCTGGACCTCAGGACGAAGCCTATGACAGGGCTCTGGCTCAC 2174
Qy 696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
   ::::::::::::::::::::|
Db 2175 CACCTGGTGCACCTGTACTACAGAGCGAGGAGCAGCAGAGAGGAGCTCTCGACATG 2234
Qy 716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
   ::::::::::::::::::::|
Db 2235 GCGGTGTAAAGACTACATTCCTACGCGCACAGCACCATCATGCGCGGCTAAGTGAG 2294
Qy 736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
   ::::::::::::::::::::|
Db 2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAG-----ATT 2342
Qy 756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
   ::::::::::::::::::::|
Db 2343 GGCAGTAGCCGGGGAATGGTTTTCGCATACCTTCGACACGCTAGAGTCATTAATCCGCTTA 2402
Qy 776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
   ::::::::::::::::::::|
Db 2403 GCAGAGGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGAAGAG 2462
Qy 796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
   ::::::::::::::::::::|
Db 2463 GCCAAACGCTCCATCGGGAAGCTCTGAAGCAGTCTGCACTGATCCCGGACTGGCATC 2522
Qy 816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
   ::::::::::::::::::::|
Db 2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGGCCACCTCTCTAAACGGAAAGAGAA 2582
Qy 836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerGly 855
   ::::::::::::::::::::|
Db 2583 TTAGCTGAAGCATTTGAAGAAGCTTATT-----TTATCTAAGGGCAAAACACCAAGCT 2633
```

```
Qy 856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
   ::::::::::::::::::::|
Db 2634 CTAAAAATACCAGCAACTTTTGAAGATATTTCGGGGCAAAATCTGACATAGCAATATAA 2693
Qy 876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
   ::::::::::::::::::::|
Db 2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGCCAGATGATGATTTCTTCACAGTGAAGTGGG 2753
Qy 896 AspIleValLys 899
   ::::::::::::::::::::|
Db 2754 AAGACCGTGGCG 2765
```

Search completed: December 6, 2005, 14:22:51
Job time : 1247 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 10:15:34 ; Search time 155 Seconds
(without alignments)

1809.092 Million cell updates/sec

Title: US-10-768-511-8

Perfect score: 4609

Sequence: 1 MENNDALDIGAVSPYPSQS.....LGSQEGFLTVGHGIVKRV 901

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3289935 seqs, 155610033 residues

Total number of hits satisfying chosen parameters: 6579870

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102336_18498/app_query.fasta_1.1095
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10768511@cgn_1_117@runat_02122005_102336_18498
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.*

1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1016	22.0	2821	7	US-11-087-227-13
2	974	21.1	2900	7	US-11-087-227-15
3	943.5	20.5	3453	7	US-11-087-227-11
4	930	20.2	2715	6	US-10-507-275-4
5	918.5	19.9	3769	7	US-11-087-227-89
6	856	18.6	3379	6	US-10-507-275-2
7	438.5	9.5	507	7	US-11-108-172-158
8	176.5	3.8	1312	6	US-10-750-185-55669
					Sequence 13, Appl
					Sequence 15, Appl
					Sequence 11, Appl
					Sequence 89, Appl
					Sequence 2, Appl
					Sequence 158, Appl
					Sequence 55669, A

ALIGNMENTS

RESULT 1

US-11-087-227-13

; Sequence 13, Application US/11087227

; Publication No. US20050260566A1

; GENERAL INFORMATION:

; APPLICANT: Fischer, Timothy J.

; APPLICANT: Malinowski, Douglas P.

; APPLICANT: Taylor, Adriann J.

; APPLICANT: Parker, Margaret R.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE

; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE

; FILE REFERENCE: 046143/287139

; CURRENT APPLICATION NUMBER: US/11/087,227

; CURRENT FILING DATE: 2005-03-23

; PRIOR APPLICATION NUMBER: 60/556,495

; PRIOR FILING DATE: 2004-03-24

; NUMBER OF SEQ ID NOS: 90

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 2821

; TYPE: DNA

; ORGANISM: Homo sapiens

US-11-087-227-13

Alignment Scores:

Score: 3.33e-62 Length: 2821

Percent Similarity: 1016.00 Matches: 298

Best Local Similarity: 46.88% Conservative: 160

Query Match: 30.50% Mismatches: 324

Indels: 195

DB: 7 Gaps: 34

651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThr 670
1981 GCCTACGGCGCTACAACTCCGCGCAGCTGGAGCAGACATACAGTACCTCTGCA 2040
671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
2041 CTGCTCTCCCGTTGACCTCTCTGGCTGATTCAGGACCGCGCCGACGACAAATGAC 2100
691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
2101 CTACGGTTGGCCCGACGACATACCTATGTGCACGACGACGCGGCGCCCTCCCGAG 2160
711 ---AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
2161 TTTGAACCTCTGGACATGAAGCTCATAGGCGCTTACATAGCCATGTGCGCGAGAAG--- 2217
730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
2218 CAGCCCATGGTCCAGAGTCTCTGGCTGACTACATCAGCAGCAGCATACGTGGAGATGAG 2277
750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
2278 CGAGAGGCT-----TGGCTAGTAGAGATGCCACCTATCTTCTGCCCGACCCCTG 2328
770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
2329 CTGGCTATCTCGCGCTTTCACCTGCTCTGGCAGCTCTGAGATGTGTGATGTGTGGAG 2388
790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
2389 AAGAAGATGTGAATGAAGCATCAGGCTAATGAGATGTCAAGGACTCTCTCTAGGA 2448
810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrGlyValSerAlaSerGlu 829
2449 GAC-----AAGGGCGCAGACAGCTAGGACT 2472
830 ArgIleArgArgAlaAsnLeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSer 849
2473 CAGAGACCGACGAGATGTGATATTTGCCACCGCTCGGTGAAGT-----GTCTCA 2520
850 ProGlySerSer-SerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
2521 GGGGGCCGAAGTGTCCGGT-----TCTCTGAGGCGACGACGAG 2556
869 rSerValAspVal-----SerLeuGlnAspIleLysAsnAlaLe 882
2557 CGCTGTGATCTCGTGGCTTCACACCCCGCAGTTCCAGGCGGCTCTGGATGAATATGAG 2616
882 uGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
2617 GAGCTCAATGTCTGGCAGGTCAATGTCTCCCGACACGCGATCATCTTTTG 2665

RESULT 2

US-11-087-227-15
; Sequence 15, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adrian J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; PRIOR FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 2900
; TYPE: DNA

; ORGANISM: Homo sapiens
US-11-087-227-15
Alignment Scores: 3,25e-59 Length: 2900
Pred. No.: 974,00 Matches: 295
Score: 47.42% Conservative: 164
Percent Similarity: 30.48% Mismatches: 344
Best Local Similarity: 21.13% Indels: 165
Query Match: 7 Gaps: 32
DB:
US-10-768-511-8 (1-901) x US-11-087-227-15 (1-2900)
QY 6 AlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGluGlyValSerThr 25
Db 121 GCGGTGTCTCTGGGTGTGATGGCAATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 180
QY 26 Pro-----LeuProGlnValThrSerProSerPheAspAsnAlaAla 39
Db 181 CCGGAGCTGTGAAGCGGCTGGAAATCCGGGCCCATGTGCTCTTTGTTTACTAAGACGGA 240
QY 40 SerPro-----ValAlaGlyArgAlaValArgGlnThrProThr 53
Db 241 AGCGATGGCGGAGCGGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 300
QY 54 SerAla-----ValArgArgArgGly-----ArgGluThr 63
Db 301 AAATCAGGGGCTAAGGGGACCCAAAGAGGCGGGGATCATAGGGGTGGAAAGAAAGCTG 360
QY 64 AspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSer 83
Db 361 AGAACCTTGAGACCGGAGTGTGAGGGGCCAACAGGGGAGGGCGCTAGAAATTTTAAACTAA 420
QY 84 ProTyrAspAlaGlyThrPro--GlyThrProGlyThrProValAlaThrPro-- 101
Db 421 AGTAGGGACCGGAATTTCCCTCGGGAGATGTTGGATGGCCCTGTGCACCTGCCACGGGCTC 480
QY 101 alTyrAlaThrProValGlyThrProMetGlyThrProSerPhe--HisArgGlyThrP 120
Db 481 TTTATTTCTTCGTGGTTAGAAACAGACTTGTGAAAAGAGTTATGCCCTTGTGGGAGA 540
QY 120 roGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 140
Db 541 CTTCCGAAAGGTTAAGAGTTCTTACAGAGTTCTACAGAGTATGATGAACTCGGGAAGAA 600
QY 140 rSerGlnSerArg-GluPro--GlyHisArgSerProSerArgGluProSerAlaAsp 158
Db 601 GCAGTTCAAGTATGGGAACAGTGTGTTGGTGGCTCATCGGGAACAGAGTGGCTCTGTA 660
QY 159 GlyArgPro-SerGluSerAlaGluProAspThrLeuGlyGlyGluTyrAlaTyrVa 178
Db 661 TGTGGACCTGGACGACGATGACCCCGAGTTG----- 700
QY 178 lTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAs 198
Db 701 -----GTGCACTCAATTTGTGAGATGCCAGCGCTACCGGAGCT 741
QY 198 nTyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleGluGluThrVa 218
Db 742 CTTTGTGTGATGCGGTACAGAGCTGTGCTCCTCAGTAC----- 778
QY 218 lGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr----- 233
Db 779 ----AAGGAGAGGGAAGTGTGTAATAAAGATGTCTCGGACGTTTACATTTAGCATCGGCT 834
QY 234 -----AspHisAspProAspLeuTyrAlaLysIleValArg----- 245
Db 835 AATGATGGACGACGCGGAGTGGGACACCTGGGATG-----GTCCGAAGACCCCA 882
QY 246 -----TyrProLeuAspIleIleProLeuLeu-----AspThr 256
Db 883 GAACCATGACCTGCTGAACTCATGCGCAGATTTGTGATGGTCTCTGTGCGGGAAGATGT 942

; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE

; FILE REFERENCE: 046143/287139

; CURRENT APPLICATION NUMBER: US/11/087,227

; CURRENT FILING DATE: 2005-03-23

; PRIOR APPLICATION NUMBER: 60/556,495

; PRIOR FILING DATE: 2004-03-24

; NUMBER OF SEQ ID NOS: 90

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 3453

; TYPE: DNA

; ORGANISM: Homo sapiens

US-11-087-227-11

Alignment Scores:
 Pred. No.: 5,848-57 Length: 3453
 Score: 943.50 Matches: 261
 Percent Similarity: 46.44% Conservative: 157
 Best Local Similarity: 29.00% Mismatches: 343
 Query Match: 20.47% Indels: 139
 DB: 7 Gaps: 23

US-10-768-511-8 (1-901) x US-11-087-227-11 (1-3453)

```

QY 31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgAlaValArgGln 50
DB 67 TCATCGGAATCTTCAACATGATCCAGCCCG---GCCAGCGTGGCGAGGCAATGAT 123
QY 51 ThrProThrSerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArg 70
DB 124 CCTCTCACTCCAGCCCTGGCGAGCTCCCGGCTAGTATGTCCTC-----ACCTCC 177
QY 71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
DB 178 AGCCCTGGCGGTGACCTT-----CCACCATTTGAGGATGATGCTCCGAG 219
QY 91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
DB 220 GGGCTCTTAGGCACA-----GAGGGGGCCCTGGAGGAGGAGGATGGAGGAGCTC 273
QY 111 -----GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGlu 127
DB 274 ATTGAGATGGCATGGAAGGAGCTACCGCGGCATCCAGAGTGGACGCCCTATGAGGCC 333
QY 128 LeuGly-----SerGlnGlyLysPro 134
DB 334 GAGGACTGGCTGTGATGATGAGGAGCTAGAGGAGCTGACGCGCAGTCAGAGGGAGGCA 393
QY 135 LeuHisArgArgArgSerGlnSerArgGluProGly----- 147
DB 394 GCAGAGCGGGCCATGCGGCGAGCTGACCGGAGGCTGGCGGGGCTGGCGCGCATGCGC 453
QY 148 -----HisArgSerProSerArgGluProSer 156
DB 454 CGTGGGCTCTGTATGACAGCGATGAGGAGGACGAGGAGCGCGCCCTCCCGCAAGCGCCG 513
QY 157 AlaAspGlyArgProSerGluSerAlaGluProAsp----- 168
DB 514 CAGGTGGAGCGGCCACGAGGAGCGGAGGAGGAGGAGGAGGATGATGAGAGCATCGAG 573
QY 169 -----AspThrLeuGlyGlyGluTyrAlaTyrValTyr-----GlyThrAsn 182
DB 574 AACCTGAGGATCTCAAGGCCACTGTGTGGCGAGTGGGTGAGTGGCGGGCCCCCGG 633
QY 183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr-----ArgSer 201
DB 634 CTG-----GAGATCCACCACCGCTTCAAGAACTTCTGCGCACTCACGTCGACAGC 684
QY 202 SerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluThrValGluArgGlu 221
DB 685 CACGGCCACCAACGTC-----TTCAAGGAGCGCATCAGCAGCATGTGCCAAAGAGAAC 735
QY 222 GluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAla 241

```

```

DB 736 CGTGAGAGCCCTGGTGGTGAACATATGAGGACTTGGCAGCAGGAGCAGCTGTGCGCTAC 795
QY 242 LysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluVal 261
DB 796 TTCTTGGCTGAGGACACCGGAGCTGTCGAGATCTTTGATGAGGCTGCCCTGGAGGTG 855
QY 262 AlaThrSerLeuLeuProThrPheGluLys-----HisIleGluAlaArgProPhe 278
DB 856 GTACTGGCCATGTACCCCAAGTACAGCCGCATCACCACCATCATCATGCTCCGCATCTCC 915
QY 279 AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuVal 298
DB 916 CACTGCTCTGCTGGAGGAGCTGCTGCTGCTGAGGAGCTGCATCTGAACCATGCTGATC 975
QY 299 SerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhe 318
DB 976 CGCACCATGGGGTGGTGACACGCTGCACCTGGCGCTGCTGCCCATCAGCTCAGCATGTG 1035
QY 319 PheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgVal 338
DB 1036 TACAACCTGCAACAGTGCATTTCTGCTGGGTCTTTCTGCGCAGTCCCAAGAACCCAGG 1095
QY 339 GluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIle 358
DB 1096 GTGAAACACCGCTCTGT-----CTGAGTGCAGTGGCGCGGCCCTTTGAGGTCAAC 1149
QY 359 HisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAla 378
DB 1150 ATGGAGGAGAGCATCATATCAGAACTTACCAGCGTATCCGAATCCAGAGAGTCCAGCAA 1209
QY 379 IleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAsp 398
DB 1210 GTGGCGGTGGCGGCTGCCCTCCAGGAGCGCATTTCTCTGCGCAGATCTGTGGGAC 1269
QY 399 AlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArg 418
DB 1270 AGCTGCAGCCAGGAGAGCAGATGAGCTGAGCTGCGATCTATCACAAC-----AAC 1320
QY 419 ValGlyProAsnGlnArgThrLeuArgAla-----LeuTyrLysThrTyrIleAspCys 436
DB 1321 TATGATGGTCCCTCAACACTGCCAATGGCTTCCCTGCTTTGCCCATCTGTCATCTCAGCC 1380
QY 437 ValHisValLysLysSerAspArg-----GlyArgLeuGlnThrGluAspPro 452
DB 1381 AACACACGTGGCCCAAGAGGACCAAGGTTGCTGTAGGGGAACTGACCGCATGAAGAT--- 1437
QY 453 MetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGlu 472
DB 1437 ----- 1437
QY 473 AlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyr 492
DB 1438 -----GTGAGATGATCACTAGCTCTCCAAGGATCAGCAGATCGGA 1479
QY 493 AspArgLeuSerArgSerLeuAlaProSerIleTyrGluLeuGluAspIleLysGly 512
DB 1480 GAGAAGATCTTTGCCAGCATCTGCTCTTCATCTATGTGTCATGAGACATCAAGAGAGGC 1539
QY 513 LeuLeuCysGlnLeuPheGlyLysAlaLysLysIleProSerGlyAlaSerPheArg 532
DB 1540 CTGGCTCTGGCCCTGTTCGGAGGGGAGCCCAAAACCCAGGTGGCAAGCAAGGTACGT 1599
QY 533 GlyAspIleAsnValLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGln 552
DB 1600 GGTGATATCAAGCTGCTTGTGGCGAGACCCCTGGCAGCAGGAGTTCGAGTTCTCAAG 1659
QY 553 TyrValHisIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaVal 572
DB 1660 TATATTGAGAAAGTGTCCAGCGGAGCATCTTCCACCATGGCCAGGCGGCGTGGCTGTG 1719
QY 573 GlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGly 592

```

```
Db 1720 GGCCTCAGCGGTATGTCCAGCGGCACCTGTGCAGCAGGAGTGGACCTTGGAGGCTGGG 1779
Qy 593 AlaleuValLeuSerAspArgGlyleCysCysIleAspGluPheAspLysMetsSerAsp 612
Db 1780 GGCCTGGTCTGGCTGACCGAGAGTGTCTCATTTGATGAATTCGACAAAGATGAATGAC 1839
Qy 613 AsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGly 632
Db 1840 CAGGACAGACAGCATCATGAGGCCATGAGCAACAGAGGNTCTCATCTCGAGGCT 1899
Qy 633 GlyIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGly 652
Db 1900 GGCATCGTCACCTCGCTGAGGCTCGTCGACGGTCATTTGCTGCCCAACCCCATAGGA 1959
Qy 653 SerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeu 672
Db 1960 GGGCGCTACGACCCCTCGCTGCTGCTTCTTGAGAACGTGGACCTCACAGAGCCCATCATC 2019
Qy 673 SerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArg 692
Db 2020 TCACGCTTTGACATCTGTGTGTGTGAGGACACCGTGACCCAGTCCAGGACGAGATG 2079
Qy 693 LeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln----- 710
Db 2080 CTGGCCCGCTTCTGTGTGGCAGCCACGTCAGACACACCACCCAGCAACAGGAGGAGGAG 2139
Qy 711 -----AspAla 712
Db 2140 GGGCTGGCCAAATGGCAGCGCTGTGAGCCCGCATGCCCAACACATATGGGTGGAGGCC 2199
Qy 713 LeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThr 732
Db 2200 CTGCCCCAGGAGTCTTGAAGAAGTACATCATCTACGCCAAGGAGGGTCCACCCGAAG 2259
Qy 733 LeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGly 752
Db 2260 CTCACACAGATGGACAGGACCAAGTGGCCAGATGTACAGTGCACCTGAGGAAGAATCT 2319
Qy 753 AsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMet 772
Db 2320 ATGGCGACAGGAGCAGC-----ATCCCATTTAGCGTGTGCGCACATCGAGTCCATG 2367
Qy 773 IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAsp 792
Db 2368 ATCCGCATGGCGAGGCCACGCGGCATCCATCTGCGGAGTATGTATGACGAGACGAC 2427
Qy 793 AlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAla 812
Db 2428 GTCACATGGCCATCCGGGTG-----ATGCTGGAGAGCTTTCATAGACACACAG 2475
Qy 813 ThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArg 832
Db 2476 AAGTTCAGCGTCATCGCAGCATGCGCAAGACT---TTTGCCCGCTAGCTTTTCATTCCGG 2532
Qy 833 ArgAlaAsn-----LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
Db 2533 COTGCAACAATAGAGCTGTGCTCTTCATACCTGAAGCAGTGTAGTGGCAGAGCAGGTGACA 2592

RESULT 4
US-10-507-275-4
; Sequence 4, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhibitory Ability
; FILE REFERENCE: 082368-001100US
```

```
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2715
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2715)
US-10-507-275-4

Alignment Scores:
Pred. No.: 3,9e-56 Length: 2715
Score: 930.00 Matches: 258
Percent Similarity: 46.85% Conservative: 160
Best Local Similarity: 28.79% Mismatches: 336
Query Match: 20.18% Indels: 142
DB: 6 Gaps: 23

US-10-768-511-8 (1-901) x US-10-507-275-4 (1-2715)
Qy 37 AsnAlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThrSerAlaVal 56
Db 25 TCAGCATCTAGCCCTGCCCGTCAGCGCGCGGATCAGTGATCCCTCCAGCCCA 84
Qy 57 ArgArgArgGlyArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeu 76
Db 85 GCGCGC-----AGCTCCAGACGTGCTGACGCCCTGACCTCCAGCCCT 126
Qy 77 GlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrPro 96
Db 127 GGCAGAGACCTC-----CCCCCATTTGAAGATGAGTCTGAGGGGCTCTGGGCACA 177
Qy 97 ValAlaThrProValTyrAlaThrProValGlyThrProMet-----GlyThrPro 113
Db 178 --GAGGGGCCCATGGAGGAAGAGATGGAGAGGAACCTCATTTGGTGTGATGGATGGAG 234
Qy 114 SerPheHisArgGlyThrProGln-----121
Db 235 AGAGACTACCGTCCCATTCGCGAGCTCGATGTCTACGAGCGCGAGGATGGCCCTGGAT 294
Qy 122 -----TyrLys 123
Db 295 GATGAAGATGTGGAGGAGCTGACAGCCNGTCAGAGAGAGGAGCTGACGGNCCATGAGG 354
Qy 124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSer-----141
Db 355 CAGCGGACCGT-----GAGGCTGGCAGAGCCCTGGGACGCATGCGCGGGGGCTGCTC 408
Qy 142 ---GlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db 409 TATGACAGCAGCAGGAAGATGAGGAGCGGCTGCCCCGTAAGCGCGCCACGCTAGAACGC 468
Qy 161 ProSerGluSerAlaGluProAsp-----AspThr 170
Db 469 GCCACAGGATGGCAGGAGGATGAAGAGATGATCGAGAGATTGAGAGATCTGGAGGAC 528
Qy 171 LeuGlyGly-----GluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
Db 529 CTCAGGGCCCACTCGGTGCGGAGCGGTGAGCATGGCAGGCGCCAGGCTG-----579
Qy 187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyr---ArgSerSerAlaHisAsp 205
Db 580 GAGATCCACCAACCGCTTCAAGAACTTCTCTGGCACCACCGTGGAGCAGCCATGGCCACAAC 639
Qy 206 LeuAsnSerLysTyrIleGlnIleLeuGluThrValGluArgGluGluAspThrLeu 225
Db 225 -----121
```


Db 640 GTC-----TTCAAGAGCGCATCAGTGNATGTGCAAGAAACCGTGAGATTG 690
Qy 236 AsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArg 245
Db 691 GTGGTAATATTAAGACCTGGCAGCCCGGAGCAGCTGTGGCATACTTCTCGCGGAA 750
Qy 246 TyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu 265
Db 751 GCACCGGCTGAGTCTGTCAGATCTTTGACAGGCTGCGCTGGAGGTGCTGTGGCCATG 810
Qy 266 LeuProThrPheGluLys-----HisIleGluAlaAargProPheAsnLeuLysAla 282
Db 811 TACCCTTAATATGACCGTATACCAACACATCCATGTGCGCATCTCCACCTGCTCTG 870
Qy 283 SerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGly 302
Db 871 GTGGAGGAGCTCGTTCACCTAGGAGCTGGACCTGACCTGACCTGATCCGTACCACTGGC 930
Qy 303 MetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeu 322
Db 931 GTGGTGACCACTGCACCGAGTCTGCGCCAGCTCAGCATGCTCAAGTACCACTGTAGC 990
Qy 323 ValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluLubProThr 342
Db 991 AAGTCAAACTTTGTACTGGGCGCTTTCTGCCAGTCTCAGAAATCAGGAGGTGAAGCTGCG 1050
Qy 343 ArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCys 362
Db 1051 TCCTGC-----CCTGAGTGCCAGTCTGCTGGGCCCTTTGAGATCAACATGAGGAGACC 1104
Qy 363 ThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGly 382
Db 1105 ATCTATCAGAACTACCAACGTATCCGATCCAGGAGAGTCCCGCAAGGTGGCGGTGGC 1164
Qy 383 GluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysPro 402
Db 1165 CGACTGCCCCCGTTCAAGAGTCCATTCCTCGCTGATCTGCTGGACAGCTGCAAGCCA 1224
Qy 403 GlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsn 422
Db 1225 GGGGAGGAGATTGAGTGCACCGCATTTACCAATAT-----AACTATGACGGCTCG 1275
Qy 423 GlnArgThrLeuArgAla-----LeuTyrLysThrTyrIleAspCysValHisValLys 440
Db 1276 CTTAACACCGCCCAACGGCTTTCCAGTCTTTGCGACTTATCTTGGCCCAACCATGTTGC 1335
Qy 441 LysSerAspArg-----GlyArgLeuGlnThrGluAspProMetGluMetAsp 456
Db 1336 AAGAAGGACAAACAAAGTAGCTGTGGGGGAGCTCACCGATGAGGAC----- 1380
Qy 457 LysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGlu 476
Db 1380 ----- 1380
Qy 477 AlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSer 496
Db 1381 -----GTGAGATGATCACCAGTCTCTCCAGGATCAGCAAAATTGGAGAGAGATCTTT 1434
Qy 497 ArgSerLeuAlaProSerIleTyrGluLeuGluAspIleLysGlyLeuLeuCysGln 516
Db 1435 GCCAGCATTTGCACCTCCATCTATGGGCATGAAGACATCAAGAGAGCGCTTGGCTTGGCC 1494
Qy 517 LeuPheGlyGlyLysAlaLysIleIleProSerGlyAlaSerPheArgGlyAspIleAsn 536
Db 1495 CTGTTTGGAGGGAGGCCCAAGAACCCAGGTGGAAAGCAACAGGTTCCAGGTGACATAAT 1554
Qy 537 ValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLys 556
Db 1555 GTGCTTTGTGGGAGCCCTGGCAGCAGCAAGTCCCAATCTCAATACATCGAGAA 1614
Qy 557 IleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAla 576
Db 1615 GTGTCTAGCCGTGCCATCTTCAACCACTGGCCAGGGTGGCTGACAGTGGGTCTCACCGCG 1674

Qy 577 TyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeu 596
Db 1675 TAGCTTCAGCGCATCCCGTTCAGAGAGTGGACCTTAGAGCGCGGAGCCCTGCTGTTCTG 1734
Qy 597 SerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSer 616
Db 1735 GCTGACCGGGGGGTGTCTCATTCAGCAGTTTGAAGATGATGACAGGACAGGAGCC 1794
Qy 617 MetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleLeuAla 636
Db 1795 AGCATCCACAGGCGCATGGAACAGCAAGCATCTCCATCTCCAGGCTGGCATCGTTACC 1854
Qy 637 SerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsn 656
Db 1855 TCGCTGCAAGCCCGTGCACCTGTATAGCTGTGTCACCAACCCCATAGGAGCGCTACGAC 1914
Qy 657 AlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAsp 676
Db 1915 CTTTCACTGACCTTCTCAGAAATGTAGACCTCAGAGGCCCATCATTTCCCGCTTTGAT 1974
Qy 677 LeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHis 696
Db 1975 GTCTGTGTGGTGGAGGACACTGTTGATCCAGTTCAGGATGAGATGCTGGCCGCTTT 2034
Qy 697 LeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAsp----- 711
Db 2035 GTGGTTGGCAGCAGCTCAGACCAACCCAGTAAACAGAGGATGAAGGTTGACTAAT 2094
Qy 712 -----AlaLeuAspLeuGln 716
Db 2095 GGTGGCACTTTGAGCCAGCCATGCCAACACATATGGCGTGAGCCCTGCTCCAGGAG 2154
Qy 717 ThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGlu 736
Db 2155 GTGCTGAAGAGTATATCATCTATGCCAAGGAGAGGCTCCGCCGAAGCTCAACCCAGATG 2214
Qy 737 AlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGly 756
Db 2215 GACCAGGATAAAGTGGCCAGGATGTACAGTGCACCTGAGGAGAGGATCCCATGGCAACGGC 2274
Qy 757 SerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSer 776
Db 2275 AGC-----ATTCCCATCAGGTGCGCCACATCGATGTCATGATCCGATGGCC 2322
Qy 777 GluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAla 796
Db 2323 GAGGCCCATGCCCGCATGCACCTCGGAGTACGTGATGGAAGACGATGTCAACATGGCC 2382
Qy 797 ValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIle 816
Db 2383 ATCCGAGTG-----ATGATGGAGAGCTTCATTGACACCCAGAACGTTTCAGCGTC 2430
Qy 817 AspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgAlaAsn--- 835
Db 2431 ATCGGAGTATCGCAAGACT---TTTGCCCGGTATCTCTCTCCGCGGAGATAACAAT 2487
Qy 836 -----LeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSer 849
Db 2488 GATCTGCTGCTTCTTACTGAGCGAGTGTGGTGGCTGAGCAGGTGACA 2535

RESULT 5

US-11-087-227-89
; Sequence 89, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139

; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89
; LENGTH: 3769
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-087-227-89

Alignment Scores:
Pred. No.: 3,84e-55 Length: 3769
Score: 918.50 Matches: 267
Percent Similarity: 45.43% Conservativity: 12.
Best Local Similarity: 31.26% Mismatches: 301
Query Match: 19.93% Indels: 165
DB: 7 Gaps: 24

US-10-768-511-8 (1-901) x US-11-087-227-89 (1-3769)

```
QY 124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgAlaArgArgSerGlnSer 143
DB 48 CGCAGCGAATCTCGGCACTAAGCAATATGACCTCGCGGCGGAGCGGCGGCG 107
QY 144 -ArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerG1 163
DB 108 CGCGAGCAGCAGCACCTCGGAGGTCCGCGACGA-----GGTGGCGCGAAGTGCC 155
QY 163 uSerAlaGluPro-AspAspThrLeuGlyGlyGlyGlyAlaTyRValTrpGlyThrAsnV 183
DB 156 AGAAACTGTTCTCGAC----- 172
QY 183 AlAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyRArgSerSerA 203
DB 173 -----TTCTTGAGGAGTTTCAGACAGC- 196
QY 203 laHisAspLeuAsnSerIlystYrIleGlnIleGluGluThrValGluArgGluA 223
DB 197 -----GATGGAGAAATTAATCTTGAATAGCAGAGAACTGATCGTCTGAGAGAA 251
QY 223 spThrLeuAsnIleAspMetSerAspIleTyRAspHisAspProAspLeuTyRAlaLysI 243
DB 252 ACACATTTGTTGAGTTTGGGACCTGGAAACAATTTAACAGCAACTTTCCACCACCA 311
QY 243 leValArgTyRProLeuAspIleIleProLeuLeuAspThrGlyCysGlnGluValAla 263
DB 312 TTCAAGAGGAGTTCTATAGAGTTTACCCTTACCTG-----TGTGG- 352
QY 263 hrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPhe----- 278
DB 353 -----GCCTTGAACAAATTCGTCAAGACCGTAAAGAGATCCCTCTTGCAAGGATTTT 407
QY 279 -----AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspI 294
DB 408 ATGTTGCATTCACAGCTGCTCCTACACACACAGATTCGAGAGCTCACCTCATCCAGAA 467
QY 294 leAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluI 314
DB 468 TTGGTTTGTCTACTCGCATCAGTGGGAGGTGGTGGGAGTCCACCAGTTCCACCAGAGC 527
QY 314 leLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrVal 334
DB 528 TTGTGAGCGGAATTTCTGTGCTTGGACTGT-----CAGACAGTGA 569
QY 334 alLysGly-----ArgValGluGluProThrArgCysGluLysProGluC 349
DB 570 TCAGGGATGTAGAACAGCAGGTTCAATACACACAGCAACATCTGCCGAATCCAGTTT 629
QY 349 ysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnI 369
DB 630 GTGCCAACAGGAGGAGTTCTTACTGGATACAAATAAATCAAGATTGTTGTTGATTTTCAA 689
```

```
QY 369 leValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValS 389
DB 690 AGGTTTCGTATTCAAGAGACCCCAAGCTGAGCTTCTCGAGGAGGATATCCCCCGCATTTAG 749
QY 389 erMetCysLeuTyRAsnThrMetValAspAlaValLysProGlyAspArgIleGluValT 409
DB 750 AAGTAATTTAAGGGCTGAAGCTGTGGATCAGCTCAAGCTGGTGACAGTGTGACTTTA 809
QY 409 hrGlyVal-----PheLysAlaMetAlaValArgValG 420
DB 810 CAGGACACACTGATTGTTGCTGCTCCTCAAGCTTAGCACACAGGAGCAGCTGCAG 869
QY 420 lyProAsnGlnArgThr-----LeuA 427
DB 870 AAACATAATCCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 929
QY 427 rgAlaLeu-----TyRlys---ThrTyRileAspCysValHisValL 440
DB 930 GGGCCTTGGTGTAGGACCTTTCTTATAGGCTGGTCTTTCTTCTGCTGCTGT----- 982
QY 440 ysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnA 460
DB 983 -----GTTGCGCCACCAACC 998
QY 460 spMetTyRAlaGlyTyRHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleG 480
DB 999 CAAGGTTTGGGGGGAAGAGCTCAGATGATGAGAACAGACAGCTGAGAGCATTAAGAACC 1058
QY 480 In-----LysLeuLysGluLeuSerLysLeuProGlyIleTyR 493
DB 1059 AAATGACTGTGAAGAATGGGAGAAAGTGTGTGATGATGATGATGATGATGATGATGATG 1118
QY 493 spArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyL 513
DB 1119 ACAATCTTTTGACAGCTGTTCTCTACTATACATGCGCAATGATGAAGTAAACCGGGTG 1178
QY 513 euLeuCysGlnLeuPheGlyClyLysAlaLysLysIleProSerGlyAlaSerPheArgG 533
DB 1179 TCGTCTGATGCTCTTGTGGTGGCTTCCAAAGACAAACAGGAGAGGGACCTCTCTTCGAG 1238
QY 533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
DB 1239 GGGACATAAATGTTTGCATTTGTGTGACCCAGTACAGTACAGTACAGTACAGTACAGT 1298
QY 553 yrValHisLysIleAlaProArgGlyIleTyRThrSerGlyArgGlySerSerAlaValG 573
DB 1299 ACGTGGAGGAGTTCAGCCCCAGAGCTGTCTACCCAGTGGTAAAGCGTCCAGTGTGCTG 1358
QY 573 lyLeuThrAlaTyRValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA 593
DB 1359 GCTTAACAGCAGCTGTTGTGAGAGATGAAGAATCTCATGAGTGTGTTGATTTAGGCTG 1418
QY 593 laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA 613
DB 1419 CTTTGATGTTGGCTGATTAATGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1478
QY 613 snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG 633
DB 1479 GGGATCAAGTTGCTATTCTATGAAGCTATGGAACAGCAGACCATATCCATCACTAAAG 1538
QY 633 lyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS 653
DB 1539 GAGTGAAGGCTACTCTGAACCGCCGAGCTCCATTTTGGCAGCAGCAACCCCAATCAG 1598
QY 653 erArgTyRAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuS 673
DB 1599 GACACTATGACAGATCAAAATCATTTGAACAGATATTAATTTGTCAGCTCCCATCATG 1658
QY 673 erArgPheAspLeuIleTyRLeuMetLeuAspLysProAspGluGlnAsnAspArgL 693
DB 1659 CCCGATTCGATCTCTCTTTATCTCTTATGATGATGATGATGATGATGATGATGATGATG 1718
```

```
QY 693 euAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaL 713
Db 1719 TTGCCAGGCGCATAGTAGATTGGCAT---TCAAGATTGAGGAATCAATTGATCGTGCT 1775
QY 713 euAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrL 733
Db 1776 ATTCCCTCGATGATATCAGAGGATATCTTCTTTTGCAGAGAG---TTTAAACCCAGA 1832
QY 733 euSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyA 753
Db 1833 TTTTCCAAAGAGTCAGAGGACTTCATTGTGGCAATATAAACATCTCCGCCAGAGATG 1892
QY 753 snPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetI 773
Db 1893 GT---TCTGGAGTGCACCAAGTCTTTCATGGAGGATTACAGTGCACAGCTTTGAGCATGA 1949
QY 773 leaArgIleSerGluAlaLeuAlaAaArgMetArgPheSerGluValValGluLysValAspA 793
Db 1950 TTGCTCTCTGTAAGCTATGGCTCGGATGCATCTGCTGTGTATGATGAGGTCCAACCTAAACATG 2009
QY 793 laAlaGluAlaValaArgLeuLeu-----AspVala 803
Db 2010 TGAAGGAAGCTTTCCGGTTACTGAATAAATCAATCATCCGTGTGGAAACACACCTGATGTCA 2069
QY 803 laLeuGlnGlnSer-----AlaThrAspHisAlaThrGlyThrIle----- 816
Db 2070 ATCTAGATCAAGAGGAAGAGATCCAGATGAGGTAGTAGAGGTGCTGTGGTGCATCAATG 2129
QY 817 -----AspMetAspLeuIleThrThrGlyValSerAla----- 827
Db 2130 GTCATGCTGACAGCCCTGCTGTGTAACGGGATCAATGGCTACATGAAGACATTAATC 2189
QY 828 -----SerGluArgIleArgAlaA 835
Db 2190 AAGAGTCTGCTCCAAAGCCTCTTAAGGTGGGCTCTCTGAGTACTGCCGAATCTCTA 2249
QY 835 snLeuLeuAlaAla---LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerS 854
Db 2250 ACCTATTGTGCTTCACTACCTCAGAAAGGTGGAAGAAGAGAG-----GACGAGT 2297
QY 854 erClyLeuLysThrSerGlnLeu-----LeuGluAspIleArgSerGlnSerS 870
Db 2298 CAGCATTAAGAGAGCGAGCTTGTAACTGCTGTAAGGAAATCGAATCAGATAG 2357
QY 870 erValAspValSerLeuGlnAspIleLysAsnAlaLeu 882
Db 2358 ACTCTGAAGAGAACATTATTAATAAAGAAATCATATA 2395

RESULT 6
US-10-507-275-2
; Sequence 2, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhibitory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3379
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31)..(2709)
US-10-507-275-2

Alignment Scores:
Pred. No.: 8,92e-51 Length: 3379
Score: 856.00 Matches: 244
Percent Similarity: 45.26% Conservative: 143
Best Local Similarity: 28.54% Mismatches: 322
Query Match: 18.57% Indels: 147
DB: 6 Gaps: 19

US-10-768-511-8 (1-901) x US-10-507-275-2 (1-3379)
QY 31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgAlaValArgGln 50
Db 13 TCATCGGAATCTTACCATGGCATCCAGCCCG---GCCAGCGTCGGCAGGCAATGAT 69
QY 51 ThrProThrSerAlaValaArgArgGlyArgGluThrAspSerAlaArgArgArg 70
Db 70 CTTCTCACTTCCAGCCCTGGCGAGCTCCCGGCTACTGATGCCCTC-----ACCTCC 123
QY 71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
Db 124 AGCCCTGGCGCTGACCTT-----CCACCATTGGAGGATGATGCCGAG 165
QY 91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
Db 166 GGCTCTCTAGGCACA-----GAGGGCCCTT-GAGGAAGAAGA 203
QY 111 GlyThrProSerPheHis-----ArgGlyThrProGlnTyrLysGlnArg 125
Db 204 GGTGAGAGAGAGCTATTGGAGATGGCATGGAAGGACTACCGGCCATCCAGAGCT 263
QY 126 SerGluLeu----- 128
Db 264 GGAGCGCTATGAGCGCGGAGCTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGC 323
QY 129 -----GlySerGlnGlyLysProLeuHis----- 136
Db 324 CAGTCGAAGAGGAGCGACAGACGCGCCATCGCGCACGGTACCGGGAGCTGGCGGGGCT 383
QY 136 ----- 136
Db 384 GGCGCGCATGCGCGTGGGCTCTCTGTATGACAGCATGAGGAGGACGAGGAGCGCCCTGCC 443
QY 137 ArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
Db 444 CGCAAGCGCGCCAGTGGAGCGCGCACGGAGGACGGCGAGGAGGACGAGCAGATGATGA 503
QY 157 AlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGlyAla 176
Db 504 GAGCATCGAGAACCTGGAGGATCTCAAAGGCCACTCTGTGCGGAGTGGGTGAGCATGGC 563
QY 177 TyrValTyrGlyThrAsnValaAsnIleProAspValLeuArgAlaIleArgArgPheLeu 196
Db 564 -----GGGCCCCCGCTGGAGATCCACCGCTTCAGAACTTCCT-----GCG 608
QY 197 HisAsnTyrArg-SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleGluG 216
Db 609 CACTCAGTCGACAGCCAGCCCAACGTC-----TTCAAGGCGCATCAGCGA 659
QY 216 uThrValGluArgGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAs 236
Db 660 CATGTGCAAAAGAGAACCGTGAGAGCCCTGGTGGTGAACACTATGAGGACTTGGCAGCAGGA 719
QY 236 pProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleProLeuLeuAspTh 256
Db 720 GCACGTGCTGGCTACTTCTCTGCTGAGGACCGCGGAGCTGCTGATGATTTTGTATGA 779
```

QY 256 rGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLys-----HisI 273
Db 780 GGCTCCCTGGAGGGTGGTACTGGCCATGATACCCCAAGTACGACCGCATCACCACCAT 839
QY 273 eGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAs 293
Db 840 CCATCTCCGATCTCCCACTGCTCTGGTGGAGGAGTGGCTGCTGGGACACTGCA 899
QY 293 pileAspLysLeuValSerValLysGlyMetValLleArgCysSerSerIleIleProG1 313
Db 900 TCTGAACAGCTGATCCGACCACTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 959
QY 313 uIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrVa 333
Db 960 GCTCAGCATGGTCAAGTACAACTGCAACAAGTGCATTTGCTGGGTCTTTCTGGCA 1019
QY 333 lValLysGlyArgValGluProThrArgCysGluLysProGluCysAlaAlaArgAs 353
Db 1020 GTCCAGAACAGGAGGTGAACAGGCTCTGT-----CCTGAGTGGCAGTGGCGCG 1073
QY 353 nAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuG1 373
Db 1074 CCCCTTTGAGGTCAACATGGAGGAGACCATCTATCAGAATACACGAGCGTATCCGAATCCA 1133
QY 373 nGluThrProAspAlaIleProGluGluThrProHisThrValSerMetCysLeuTy 393
Db 1134 GGAGAGTCCAGCAAGTGGCGGCTGGCGGCTGGCGGCTCCCAAGGACGCCATTCCTCT 1193
QY 393 rAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLy 413
Db 1194 CGCATCTGTGGAGCAGCTGCAACCGCAGGAGACGAGTAGAGTACGTGGCATCTATCA 1253
QY 413 sAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAla-----LeuTyxly 431
Db 1254 CAAC-----AATATGATGGTCCCTCAACACTGCCAATGGCTTCCCTGTCTTGGC 1304
QY 431 eThrTyrlleAspCysValHisValLysLysSerAspArg-----GlyArgLe 447
Db 1305 CACTGTCTATCTTAGCCCAACCGCTGGCCCAAGAGGACCAACAGTGTGTAGGGGAACT 1364
QY 447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisG1 467
Db 1365 GACCATGAAGAT----- 1377
QY 467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerIly 487
Db 1378 -----GTGAAGATGATCACTAGCTCTCCAA 1403
QY 487 sLeuProGlyIleTyAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuG1 507
Db 1404 GGATCAGCAGATCGGAGAGAGATCTTTGGCCAGCATTTGCTCTCCATCTATGTGTATGA 1463
QY 507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSe 527
Db 1464 AGACATCAAGAGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1523
QY 527 rGlyAlaSerPheArgLysAspIleAsnValLeuLeuValGlyAspProGlyThrSerIly 547
Db 1524 CAAGCACAAGGTACGTGGTGTATCAACGTGCTCTTGTGCGGAGACCTTGGCCACAGCA 1583
QY 547 sSerGlnLeuLeuGlnTyValHisLysIleAlaProArgGlyIleTyThrSerGlyAr 567
Db 1584 GTCCGAGTTTCTCAAGATATATTGAGAAGGTGCCAGCCGAGCATCTTCACCACTGGCCA 1643
QY 567 gGlySerSerAlaValGlyLeuThrAlaTyValThrLysAspProGluThrArgGluTh 587
Db 1644 GGGGGCGTGGCTGTGGCGGTATGTCAGCGCGACCTGTGTCAGGGAGGTG 1703
QY 587 rValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysValIleAspGluPh 607
Db 1704 GACCTTGGAGGTGGGGCCCTGTTCTGGCTACCGGAGGAGTGTCTCATTTGATGATTT 1763
QY 607 eAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnThrVa 627

Db 1764 TGACAGATGATGACCAAGGACAGACAGCATCCATGAGCCATGGAGCACACAGCAT 1823
QY 627 lSerValAlaLysGlyGlyIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCy 647
Db 1824 CTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGAGGCTCGTGCACGGTCAATTGCTGC 1883
QY 647 sAlaAsnProSerGlySerArgTyAsnAlaArgLeuSerValLleAspAsnIleGlnIe 667
Db 1884 CGCCAACCCCATAGGAGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAAGCTGGACCT 1943
QY 667 uProProThrLeuLeuSerArgPheAspLeuIleTyThrMetLeuAspLysProAspG1 687
Db 1944 CACAGAGCCCATCATCTCAGCTTTGACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2003
QY 687 uGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyArgLeuTyGluVa 707
Db 2004 AGTCAGGACGAGATGCTGGCCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2063
QY 707 lSerLysGln----- 710
Db 2064 CAACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2123
QY 711 -----AspAlaLeuAspLeuGlnThrLeuThrAlaTyrlleThrTyrAlaArgL 727
Db 2124 GTATGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2183
QY 727 nHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValG1 747
Db 2184 GAGGTCCACCGGAGCTCAACAGATGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2243
QY 747 uMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProAr 767
Db 2244 CCTGAGGAAGAATCTATGCGCAGCAGGAGC-----ATCCCATTTACGCTGCG 2291
QY 767 gGlnLeuLysSerMet-IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluV 787
Db 2292 GCACATCAGTCCCATGAGTC---ATGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2348
QY 787 alValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db 2349 ATGTGATCGAAGACGAGCTCAACATGGCCATCCGCGTGATG 2389
RESULT 7
US-11-108-172-158
; Sequence 158, Application US/11108172
; Publication No. US20050260177A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongcong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSTICS
; FILE REFERENCE: 210121.471C15
; CURRENT APPLICATION NUMBER: US/11/108,172
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 10/025,380
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 09/922,217
; PRIOR FILING DATE: 2001-08-03


```

; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45476
; LENGTH: 1851
; TYPE: DNA
; ORGANISM: Bovine
; ORGANISM: Bovine 19866880565574
US-10-750-185-45476

Alignment Scores:
Pred. No.: 0.00817 Length: 1851
Score: 174.00 Matches: 34
Percent Similarity: 74.60% Conservative: 13
Best Local Similarity: 53.97% Mismatches: 14
Query Match: 3.78% Indels: 2
DB: 6 Gaps: 1

US-10-768-511-8 (1-901) x US-10-750-185-45476 (1-1851)
QY 571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
DB 1366 TCCTTAGTCTGACGGTAAGTCTGTAAGAGAC-----TCAGGAGAAATGGAAATTTGGAG 1419
QY 591 SerGlyAlaLeuValLeuSerAspArgGlyLeuCysCysIleAspGluPheAspLysMet 610
DB 1420 GCGGAGGAGATAGTCTCCGACAGATCTGCGCTCTGCTATGCAATTTAATAGCCTC 1479
QY 611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630
DB 1480 AAAGAGCATGACAGAACTAGTATCCACGAGCAATGGAGCAAAACCATCAGTGTGCT 1539
QY 631 LysGlyGly 633
DB 1540 AAGGCTGGG 1548

RESULT 10
US-11-004-057-3
; Sequence 3, Application US/11004057
; Publication No. US20050244846A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Gary L.
; TITLE OF INVENTION: MEK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING
; FILE REFERENCE: APOPTOSIS
; CURRENT APPLICATION NUMBER: US/11/004,057
; CURRENT FILING DATE: 2004-12-02
; PRIOR APPLICATION NUMBER: US/09/403,075
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 5253
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(4496)
US-11-004-057-3

Alignment Scores:
Pred. No.: 1.52 Length: 5253
Score: 150.00 Matches: 224
Percent Similarity: 32.62% Conservative: 125
Best Local Similarity: 20.93% Mismatches: 364
Query Match: 3.25% Indels: 358
DB: 9 Gaps: 56

US-10-768-511-8 (1-901) x US-11-004-057-3 (1-5253)
QY 11 AlaValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnVal 30
DB ||||| ||||| ::||| |||||

```

Qy	359	HisAsnArg-CysThrPheAlaLeuLys-----GlnIleValArgLeuGlnGluThr	375
Db	936	AGAGCCAGGCTGACCTGCTGCAGCAGATAGGACCACTCTTCTCTGATTGGAGAGAC	995
Qy	375	rProAspAlaIleProGluGly-----GluThrPr	385
Db	996	AGTCCAGACAATAATACCCGGTGTATTATGGGCCACAGAACTGCAGCTGTGGCGCTGGA	1055
Qy	385	oHisThrVal---SerMetCysLeuTyraAsnThrMetValAspAlaValIleProGlyAs	404
Db	1056	GCATTCTGTATTACACTCTTGTGTGTC-ARGCTCCGGGTGTTTCAGCTAGAACCCCTCTCA	1114
Qy	404	pArgIleGluValThrGlyValPheIysAlaMetAlaValArgValGlyProAsnGlnAr	424
Db	1115	CCCATGTTATGGAGAAACTTTAAATAATTTTCAGGTT-----	1154
Qy	424	gThrLeuArgAlaLeuTyrlsThrTyrlleAspCysValHisValIlysLysSerAspAr	444
Db	1155	-----GAGAGTTTGTTCAGAAATAC-----CACAGTAGGCGTAGCTCGAG	1195
Qy	444	gGlyArgLeuGlnThrGluAspProMetGlu-----MetAspLysGluAs	459
Db	1196	AATCAAAGCTCCATCCCGGAACACCATCCAGAAAGTTGTGTCAAGCATGTCAAAATCTCA	1255
Qy	459	nAspMetTyAlaGlyTyrlHisGluSerAspThrSerGluAlaAlaAsnGluAlaIysIl	479
Db	1256	CACACTG-----TCATCGTCTAGCACATCCATCATCTAGTTTTCAGAAACAGCAT	1303
Qy	479	eGlnLysLeuLysGluLeuSer-----LysLeuProGlyIleTyraAspArgLeuSe	496
Db	1304	CAAGATGAAGAGAGCAGATGTGTCCCATCTGCTGTGGCATGCTGTGATCAGGAGAG	1363
Qy	496	rArgSerLeuAlaPro-----SerIleTr	504
Db	1364	CCTGACTGTGTGTGAAGATGGCTGCAGGAACAAGCTGCACCACTGTCATGTCCATCTG	1423
Qy	504	p-----GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyLys--	521
Db	1424	GGCGGAAGAGGTAGAGAAATAGAGAGCCCTTAATATGTCCCTTTTGTAGATCTAAGTG	1483
Qy	522	-----AlaLysLysIleProSerGlyAlaSerPh	531
Db	1484	GAGATCCCATGACTTCTACAGCCATGAGTTATCAAGCCCGTGAGGTCCCCCGCTCCCT	1543
Qy	531	eArg-----	532
Db	1544	GCGAGCTGTCCAGCAGCCATCTCTCCCGCAGCAGCCCGTGGCCGATCAAGCGGAGAA	1603
Qy	533	-----GlyAspIleAsnValLeuValGlyAsp-----ProGlyThrSerLy	547
Db	1604	TCAGGAGAGCAGTTTAACCTTACTCATTTTGGAAACCCAGCAGATTCTTCCGCTTACAA	1663
Qy	547	sSer-----	548
Db	1664	AGATTGGCCGAGCATGGATTCAAGTGTTTTGGAACTGGAACTCGTGGCTGCTTATTCTC	1723
Qy	549	-----GlnLeuLeuGlnTyraValHisLysIleAla-----	558
Db	1724	TAGAAACTGGAAACGTAAAGGAAATGGCCCTTAGCGCTTTTCCACGACGTATTAGTGGGC	1783
Qy	559	-----ProArgGlyIleTyThrSerGlyArgGlySerSerAl	571
Db	1784	CCTGTTGTGGCAACCGGGGAGAGCATTGGAAACTCTGGAGCGCGCAGTGGGGCAGCTT	1843
Qy	571	aValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSe	591
Db	1844	AAGCGCGGAGCGCCAGCGGGTCTCCACGCCAGCATCTCAGGGATGTGTGGAGGC	1903
Qy	591	rGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSe	611
Db	1904	GTGCTGCAGTGTCTCTCT-----ATAGTCTGGCTGACCTGTCTACAAGTGTA	1954

611 rAspAnla-----ArgSerMetLeu-----HisGluValMe 622
1955 CGTGTGCTGCTTTAAAAACATTCAGAGCCATGCTGGTATATACACTCTTGTGCACAGTCTGC 2014
622 tGluGln-----GlnThrValSerValAlaLal 631
2015 AGAAAGANTCAACATTCAGAGACTCTCCGGCCAGTTGTAGACACTATCTTGTCAAG-- 2072
631 sGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAla- 646
2073 -----TGTGCAGATGCCAACACGCGCACAGTGCAGCTGCTCCATATCTACAGTGCT 2122
647 -----CysAlaAsnProSerGlySerArgTyrAsnAlaAaG- 658
2123 GGAACCTCTGCAAGGCCCAAGCAGGAGAGCTGGCGGTGGGAGAGAAATACTTAAAGCTGG 2182
659 -----LeuSerValIle- 662
2183 GTCCATCGGGTGTGGTGTGATTAAGTCTTAAAGTTGTATCTTGGAAACCAAGCTGA 2242
663 -----AspAsnIleGlnLeuProThrLeuLeuSerArgPheAspLeuIle---TyrLe 680
2243 ATCAAAACAACGTGCAA-----GAACTGCTGGGTGCGCTCTGCTTTATAGACAGGTT 2293
680 uMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLe 700
2294 GCTGTTGGAAATTCCTGCTGAA-----TTCTATCTCATATTGTCAGTAC 2338
700 uHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu-----GlnThrLe 718
2339 T-----GATGCTCACAAGCTGAGCCCTGTTGAAATCAGGTACAGGAAGCT 2383
718 uThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAl 738
2384 GCTCTCCCTCTTAACCTTTGCTTGCATCCATGCAATTCACATTCGCCATCG----- 2432
738 aGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSe 758
2433 -----ATGGTGGCAAGCTCTCTCGAGG-----ATATATCTGAGCTC 2470
758 rLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAl 778
2471 TGCCAGGATGTCACCGCAGTGCCTGCTGTGTTTTCACAGCTGTAACCATGCTTATATGC 2530
778 aLeuAla-----ArgMetArgPheSerGluValValGluLysVa 791
2531 TTCTGGCTCCACCCACTTCACAGGATGCCCGCGCTGATGGCTATCCGGATGAGGT 2590
791 lAspAlaAlaGluAlaValArgLeu-----LeuAsp----- 801
2591 AGAAATGCCAGGTATCCAGCTGGGTGGAGGACACTGTGGATGGGCATCAGGACAG 2650
802 -----ValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGl 814
2651 CTTACAGGCGGTGGCCCCCACCAGCTGCTCTAGAAAAACAGCTCCCTTCAGCAC----- 2702
814 yThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgAl 834
2703 -ACAGTCCATAGAGAGAAAACTGAAAAGGACTTAAGTGCTACGAGA----- 2747
834 aAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSe 854
2748 -----CTGAGTGCCAGCTCGGAGGACATTTCTGACAGACTGGCC-----GGCGTCTCTGT 2797
854 rGlyLeuLysThrSerGlnLeuLeuGlu 863
2798 AGGACTTCCCAGCTCAACAAACACAGAA 2825
Dp

RESULT 11
US-10-750-185-27656
; Sequence 27656, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 27656
 ; LENGTH: 876
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866881490050
 ; US-10-750-185-27656
 Alignment Scores:
 Pred. No.: 0.425 Length: 876
 Score: 144.00 Matches: 47
 Percent Similarity: 28.69% Conservative: 23
 Best Local Similarity: 19.26% Mismatches: 23
 Query Match: 3.12% Indels: 151
 DB: 6 Gaps: 4
 US-10-768-511-8 (1-901) x US-10-750-185-27656 (1-876)
 Qy 324 CysGlyHisSerProLeuValThrValValValGlyValGluGluProThrArg 343
 Db 3 TGTGCCACAGC---GCCCGCTGGAAATAGACCGTGGCGGATTCGCCGAGCCCTCGGTG 59
 Qy 344 CysGlyHisProGluCysAlaAlaArgAenAlaMetSerLeuLeuHisAenArgCysThr 363
 Db 60 TGCAGGCGC---TGCCATACAGCCACAGCAGTGGCCTCATCCACACCGCTCTGTG 113
 Qy 364 PheAlaAenLysGlnIleValArg----- 371
 Db 114 TTTTCGACAGACAGATGGTACGTGCCCCCTCCACGTGCCCTGTCTCTGTCGCCCTTG 173
 Qy 371 ----- 371
 Db 174 CCGGGCCCTCTGAAGCCCAAGTGTAACTGCCCACTCATGTAACTTAAACCCGCGTTCC 233
 Qy 371 ----- 371
 Db 234 TGACTGGTTGACCTGGAAGGAGTACCGTCCAGATGAATTGTAAGACAGGAGAGGG 293
 Qy 371 ----- 371
 Db 294 TGATCAGGCTACCCACCGCCACCTGTATGTGGCCACCCACATAGGGGAGCATTAATC 353
 Qy 371 ----- 371
 Db 354 AACATGCTCTGTAATGAGAGGTAGACTGAGTGTGCTGGAGACTAGTGTGTGTGTGAG 413
 Qy 371 ----- 371
 Db 414 CATAGCCATCTCCAGAACAACTCCCTGGGCGAGTGGAGTGCCAGGTATCTGGATGTC 473
 Qy 372 -----LeuGlnGluThrProAepAlaIlePr 380
 Db 474 CTGAGATGTCCATTCGCCCACTCTCCAGATCAAGCTCCAGGAGTCCCTCTGAAGACATGCC 533
 Qy 380 oGluGlyGluThrProHisThrValSerMetCysLeuTyrAenThrMetValAspAlaVa 400
 Db 534 CGGGGGCAGACCGCCAGCGGTGCTCTTTGCTCACAATACCTCTGTCGAGCAAGT 593
 Qy 400 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGl 420


```
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-11-077-716-1

Alignment Scores:
Pred. No.: 127 Length: 33014
Score: 137.00 Matches: 13.
Percent Similarity: 35.59% Conservative: 74
Best Local Similarity: 22.74% Mismatches: 250
Query Match: 2.97% Indels: 125
DB: 7 Gaps: 20

US-10-768-511-8 (1-901) x US-11-077-716-1 (1-33014)

QY 37 AsnAlaIaSerProValaIaGly-----ArgArgAlaValaArgGlnThrPro 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7995 CACGAGAGGCGCCAGATTACAGAGATCAACCGCGCGTCTGCGTTCGCCAGCAACT 7936

QY 53 ThrSerAlaValaArgArgGlyArgGlnThrAspSerAlaIaIqArgArgSerArg 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7935 ACGCGGAGACACAGCTCTCCGCC-----GCCGCGCGGAGCTGCCCT 7888

QY 73 SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThr 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7887 GCCCCTCTCCCGCGGGTCC-----GGAGCCCGCCCTACC 7852

QY 93 ProGlyThrProValaIaThrProValTyrAlaThrProValGlyThrProMetGlyThr 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7851 TCCGGGGCTCGCCGCGCTCACCGCTTTTAGAT-----GCATCATCAAGGACACC 7801

QY 113 ProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGly 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7800 CCCGGGCCACCGCCC-GCCGCGGTACCTAGTCCGCC-----GCCGGGATG 7751

QY 133 LysProLeuHisArgArgArgSerGlnSerArgGluProGlyHis-----148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7750 CGGCCTCTTGCAAGCATCGACGCCGCCACCAACCGCCCTTGAAATTAGGTATCACT 7691

QY 149 ---ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluPro 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7690 GGATCTAGCCCGCCCTGACCCGTCTATCGAGGTAAACCTGCAGGAGCTCCGCCCTGA 7631

QY 168 AspAspThrLeuGlyGlyGlnTyrAlaTyrValTyrGlyThrAsnValAsnIleProAsp 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7630 CCTGACGCCGGGAGCT-----CCAGAC 7607

QY 188 -Val-----LeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerAl 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7606 CATGACAGCTCCCATCTGCGGATGTTGTCATCAAGCTCCGACCGCGCGCGGACAT 7547

QY 203 AhisAspLeuAsnSerLysTyrIleGlnIleleGluThrValGluArgGluAs 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7546 CTGGACTTTGGCTCGCGCGGTGTGTGTCGATCCACCGTAACCTCCCTCGAGCAGCC 7487

QY 223 pThrLeuAsnIleAspMetAspIleTyrAspHisAspProAspLeuTyrAlaLysI1 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7486 AGACGCTCAGGACAGCCGGAAGTAGAAGACCACCAAGCCCAACCCCGCCAGCGGAGG 7427

QY 243 eValArgTyrProLeu-----AspI1 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7426 GCTCAATTCCTCACTCTGCTTCTTGTGCGCGGTCTCAGGTCAACCTCGTCAGGATG 7367

QY 250 eIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGl 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7366 ACAGCCCGTG-----CACCGCTGCGAGTACTGCGCAGCGTTTTCACAAAAGCCAGCAGGA 7313

QY 270 uLysHisIleGluAlaArgProPhe-----AsnLeuLysAlaSerValHi 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7312 GTGTTCGGCCCTCGCAGGACTTCTACTTCCACCATCAATAGCCACTCTCTCAATTG 7253

QY 285 sMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValI1 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7252 GTGGCGGAGATCCAGTTCTTCCCGATCGCTCGCATCTCCGACCGAGCGTCTTTGT 7193
```

```
QY 305 eArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuVal----- 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7192 CACCTACGATGATAGACCTATATCTTGATGGGGCCCTTTGGAGAGCAGCTCGTCCCTT 7133

QY 324 -----CysGlyHisSerProProLeuValThrValValLysGlyArgVa 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7132 CATGCTGTCATGAAGTTCCGCGGAGATGACCTCTAGTACTGCGCGCGAGACCT-AG 7074

QY 338 lGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuI1 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7073 CCGCAACCTTGG-----ATGGGACCGCTCGGAACAAGACCCGCTTACTTCTACT 7023

QY 358 eHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAl 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7022 GCATCCCCAGAAAAAATGGCCATAGTCCCGAGTTTAGGACCTTTCGCGACCACTGC 6963

QY 378 aIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAs 398
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6962 AAATGCTAATGCCCGTCACT-----6941

QY 398 pAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValAr 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6940 -----GTGGAGCTCATTCGTCGCTTCCAAACCTCATCTTGCAG 6903

QY 418 qValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHi 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6902 ACTGGGCGCTTT-CAGAGACGCGGTCACT-----CC 6871

QY 438 sValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGl 458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6870 CTTGAGAGCTCAGCTACGAGGAATTAATAAATTCCTTCCATCAAGGCGATCCGCGC 6811

QY 458 uAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLy 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6810 TTTCTTGAACCTTATTCATTTGTGGGCCACAAACATCAACGCTTTGACGAGATCGTG-CTGCG 6752

QY 478 sIleGlnLysLeuLysGluLeuSerLysLeuProGlyLysLeuTyrAspArgLeuSerArgSe 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6751 CGCCCAAGTAATTAACAAACCGTTCGAGTCCGCGGACCCCTTC---CGCATCACACGAA 6695

QY 498 rLeuAlaProSerIleTrpGluLeu-GluAspIleLysLysGlyLeuLeuCysGlnLeuP 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6694 CTTTATGCTCGCGCGGAAAGATACCTTTCACAGATGTCACCTTCGCCCTGCCAAATC- 6636

QY 518 heGlyGlyAlaLysLysIleProSerGlyAla-----SerPheArgG 533
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6635 -----CGCGTTCMAAAGCGCAGCGACTTTTGTCTTGGGAGCAGGG 6593

QY 533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6592 CGGATGCGACAGACA---CTGACTTCAATATACAGTACCTCAAGTCAAGTCTGGTCAGGACAC 6536

QY 553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6535 CTTTGGCTCACCCACCTCGC-----TCCGGAAGG 6504
```

RESULT 14

```
US-10-623-155-479/c
; Sequence 479, Application US/10623155
; Publication No. US20050261166A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Peckham, David W.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C20
; CURRENT APPLICATION NUMBER: US/10/623.155
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: FastSeq for Windows Version 4.0
```

; SEQ ID NO 479

; LENGTH: 34555

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 34543,34544,34545,34546,34547

; OTHER INFORMATION: n = A,T,C or G

US-10-623-155-479

Alignment Scores:

Pred. No.: 135 Length: 34555

Score: 137.00 Matches: 131

Percent Similarity: 35.5% Conservative: 74

Best Local Similarity: 22.74% Mismatches: 250

Query Match: 2.97% Indels: 125

DB: 6 Gaps: 20

US-10-768-511-8 (1-901) x US-10-623-155-479 (1-34555)

QY 37 AsnAlaSerProValalaGly-----ArgArgAlaValArgGlnThrPro 52

Db 10022 CACGAGAGGCCAGATTCCAGAGATCAACCGCGTGTGCGTTCGCCAGCAACT 9963

QY 53 ThrSerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArgSerArg 72

Db 9962 ACGCGCGCAGCACCACTCTCCCGC-----GCCGCGCGCGACGTGCCCT 9915

QY 73 SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThr 92

Db 9914 GCCCCTCTCCCGCGGTCC-----GGAGCCCGCCCTACC 9879

QY 93 ProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThr 112

Db 9878 TCCGGGGTCCCGCGGTCCAGCGTTCAGT-----GCATCATCCAAGGACACC 9828

QY 113 ProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGly 132

Db 9827 CCGCGGCCCCACCGCC-CGCCGCGGTACTCGTGTGCGGC-----GCCGCGGTG 9778

QY 133 LysProLeuHisArgArgArgSerGlnSerArgGluProProGlyHis-----148

Db 9777 CGGCTCTTGCAAGCCATCGACGCCGCCACCAACCGCCCTGGAAATTAGGTATCACT 9718

QY 149 ---ArgSerProSerArgGluProSerAlaAspGlyArgProSerAlaGluPro 167

Db 9717 GGATCTAGCCCGCGCTGACCGCTCTATCGAGGTAACTCGACGAGCTCCCGCTGA 9658

QY 168 AspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGlyThrAsnValAsnIleProAsp 187

Db 9657 CTGACCGCGCGGAGCT-----CCAGAC 9634

QY 188 -Val-----LeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAl 203

Db 9633 CATGACAGCTCCCATCTGCGGATGTGTGTCATCAAGCTCCGACCGCGCGCGGACAT 9574

QY 203 aHisAspLeuAsnSerLysTyrIleGlnIleLeuGluThrValGluArgGluAs 223

Db 9573 CTGGACTTTGGGTGCGCGCGTGTGTGTCGATCCACCGTAACTCCCTCCGACGAGCC 9514

QY 223 pThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLys11 243

Db 9513 AGACGTCAAGCAAGCAGCCGAGTAGAAGACCAACCGCCAGCCAGCGGAGGG 9454

QY 243 eValArgTyrProLeu-----Asp11 250

Db 9453 GCTCAATTCCCACTCTGCTTCTGTGCGGTGTCAGGTCAACTCGTGCGAGATGT 9394

QY 250 elleProLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheG1 270

Db 9393 ACAGCCCGTG-----CACCGCTGCCAGTACTGCGCAGCGTTTTTACAAAAGCCAGCACA 9340

QY 270 uLysHisIleGluAlaArgProPhe-----AsnLeuLysAlaSerValHi 285

Db 9339 GTGTTCCGCCGTGCGAGGACTTCTACTTTCCACCATCAATAGCCACTCTCTCAATTG 9280

QY 285 sMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetVal11 305

Db 9279 GTGGCGGAGATCCAGTTCTTCCGATCGGTCTCGCATCTCGACCGAGCGTCTCTTTGT 9220

QY 305 eArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuVal-----323

Db 9219 CACCTACATGTAGAGACTATATCTGATGGGGCCCTTTGGGAAGCAGCTCGTCCCTT 9160

QY 324 -----CysGlyHisSerProProLeuValThrValValLysGlyArgVa 338

Db 9159 CATGCTGTTCATGAAGTTCCGCGGAGATGAGCTCTAGTACTGCGCGGAGACCT-AG 9101

QY 338 lGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu11 358

Db 9100 CCGGAGACCTTGG-----ATGGGACCGCTGGGAACAAGACCGCTTACTCTTACT 9050

QY 358 eHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAl 378

Db 9049 GCATCACCACCCAGAAAAATGCGCATAGGTGCGCAGTTTAGGACCTTTCGCGACCTGC 8990

QY 378 aIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAs 398

Db 8989 AAATGCTAATGGCCCGTCACT-----8968

QY 398 pAlaValLysProGlyAspArgGlyGluValThrGlyValPheLysAlaMetAlaValAr 418

Db 8967 -----GTGGAGCTCATTCGTGCTTCCCAACCTCATCTTGGCAG 8930

QY 418 gValGlyProGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHi 438

Db 8929 ACTGGGCGCTTT-CAGACGCGGCTCAGCT-----CC 8898

QY 438 eValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysG1 458

Db 8897 CTTGAAGAGCTCACTACGAGNACTTAAATAATTCCTTCCATCAAGGCGCATCCGCGC 8838

QY 458 uAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaL 478

Db 8837 TTTCTTGGAACTTTACATTGTGGGCCACCAACATCAACGGCTTTGACGAGATCGTG-CTCGC 8779

QY 478 sIleGlnLysLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSe 498

Db 8778 CGCCCGAGGTAAATAACACCGTTCCGAGTCCGCGGACCCCTTC---CGCATCACGCAA 8722

QY 498 rLeuAlaProSerIleTyrGluLeu-GluAspIleLysLysGlyLeuLeuCysGlnLeuP 518

Db 8721 CTTTATGCTCGCGCGGGAAGATATCTTCAACGATGTCACCTTGGCCCTGCCAATC- 8663

QY 518 heGlyGlyLysAlaLysIleProSerGlyAla-----SerPheArgG 533

Db 8662 -----CGCGTTCCAAAAGCGCAGCTTTTCTCTGGGAGCAGGG 8620

QY 533 lYAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553

Db 8619 CGAGTCCGAGACA---CTGACTTCAATACCACTACCTCAAAGTCATGTCAGGAGCAC 8563

QY 553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567

Db 8562 CTTTGGCTCACCCACACCTCGC-----TCCGGAAGG 8531

RESULT 15

US-11-074-176-47

; Sequence 47, Application US/11074176

; Publication No. US20050250135A1

; GENERAL INFORMATION:

; APPLICANT: Klaenhammer, Todd R.

; APPLICANT: Russell, William M.

; APPLICANT: Altermann, Eric

; APPLICANT: McAuliffe, Olivia

; APPLICANT: Perill, Andrea Azcarate

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 09:16:54 ; Search time 5712 Seconds
(without alignments)
7380.105 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALDIGAVSPYPSQS.....LGSIQGRGFLVHGDIKRV 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODSL=frame+_p2n.model -DEV=xlh
-Q=/cgm2_1/USPTO.spool/US10768511/runat_02122005_102334_18419/app_query.fasta_1.1095
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511@cgm_1_4289/runat_02122005_102334_18419 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -YGAPOP=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2474	53.7	2699	CNS0A83A	BX820670 Arabidops
2	2337.5	50.7	3168	10 CL959068	OsIFCC002
3	1822.5	39.5	3301	4 AK088796	Mus muscu
4	1818.5	39.5	3285	4 AK089999	Mus muscu
5	1809.5	39.3	3545	4 AK036653	Mus muscu
6	1808.5	39.2	3316	4 AK011743	Mus muscu
7	1803.5	39.1	2522	10 AY408033	Homo sapi

8	1799.5	39.0	2519	10 AY408035	AY408035 Mus muscu
9	1661	36.0	2357	10 AY408034	AY408034 Pan trogl
10	1334	28.9	1607	4 CR598188	full-leng
11	1300	28.2	3389	4 DO057540	Antonopso
12	1252	27.2	1833	4 CR606514	full-leng
13	1191	25.8	879	8 DR495517	WS0298 BR
14	1165	25.3	881	7 CO123277	GR_EB05F
15	1127	24.5	1156	7 CK210237	FGA502203
16	1097	23.8	970	8 DR503163	EST436277
17	1080	23.4	707	2 BF051102	WS0298 BR
18	1060	23.0	788	3 BM412106	EST586433
19	1039	22.5	775	3 BM410852	EST585179
20	1037	22.5	1082	5 BX836107	EX836107
21	1024	22.2	709	8 DR917805	EST110934
22	1020	22.1	750	6 CA258367	SCCCT300
23	1004	21.8	755	8 DN232047	MEST1067
24	997.5	21.6	1133	5 BX363316	BX363316
25	985	21.4	978	7 CK179662	EST768982
26	978.5	21.2	1109	7 CK027527	AGENCOURT
27	975.5	21.2	798	6 CF449139	EST685484
28	973.5	21.1	765	6 CF442356	EST678701
29	966	21.0	2387	4 BC023232	Mus muscu
30	964.5	20.9	986	11 CNS06P5P	AL408947 T7 end of
31	962	20.9	943	8 DN561478	DN561478
32	954.5	20.7	727	7 CK072837	6803681c
33	951.5	20.6	1142	3 BM479183	AGENCOURT
34	947.5	20.6	3295	4 AK037751	Mus muscu
35	943.5	20.5	2709	10 AY418605	AY418605 Homo sapi
36	940.5	20.4	3283	4 AK088156	Mus muscu
37	939	20.4	2205	10 AY420810	AY420810 Homo sapi
38	939	20.4	2452	4 CR591959	full-leng
39	939	20.4	2469	4 CR611599	full-leng
40	939	20.4	2508	4 CR615349	full-leng
41	936	20.3	2709	10 AY418607	Mus muscu
42	935	20.3	725	6 CA258377	SCCCT300
43	933.5	20.2	1825	4 CR605715	full-leng
44	929.5	20.2	3177	10 CL980638	CL980638 OsIFCC035
45	927.5	20.1	2480	4 CR601762	full-leng

ALIGNMENTS

CNS0A83A 2699 bp mRNA linear HTC 06-FEB-2004
Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLRPH582F05 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).

Accession BX820670.1 GI:42467030

Version HTG, GSLT cDNA.

Source Arabidopsis thaliana (thale cress)

Organism Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 2699)

Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C.,

Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,

Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.

Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:

A Combined Approach to Evaluate and Improve Arabidopsis Genome

Annotation

Unpublished

2 (bases 1 to 2699)

Genoscope.

Direct Submission

Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

The sequences are based on single pass reads.

Life Technologies (a division of Invitrogen) members carried out

full-length libraries construction : Temple G.


```
Db 1525 ACAAGTGGCGGAGGAGCTCAGCTGTGGTTTAAACAGCTTATGTAGCTAAAGATCCTGAG 1584
Qy 584 ThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCys 603
Db 1585 ACAGAGAAACTGTTTGGAGAGTGAGCTCTTGTTCTTAGTGACGAGGTATCTGCTGT 1644
Qy 604 IleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGlu 623
Db 1645 ATTGATGAATTTGACAAATGTCGACAGTGGCGAGGAGCATGCTACACGAGGTGATGAA 1704
Qy 624 GlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSer 643
Db 1705 CAGCAGACTGTTTCAATGACAAAGCTGGTATTATTGTCATCTCTAAATGCCAGAACCTCT 1764
Qy 644 ValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAsp 663
Db 1765 GTGTTGGCTTGTGCAATCTAGTGGCTCAGCTTATTAATCCCGGCTTCTGTTATTGAG 1824
Qy 664 AsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAsp 683
Db 1825 AATATTACCTTCTCCCAACCTTGCTATCTAGATTTCGATTGATCTACTTGTATTGAC 1884
Qy 684 LysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGlu 703
Db 1885 AAGCCGATGACGAGACTGACCGAAGCTTGCAAAAGCATATTGTGGCCCTTCACCTTGAG 1944
Qy 704 AsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuAlaTyrIleThr 723
Db 1945 AACGTGAGAGTGCTCAGGAGGAGCTATAGATATTATCTACTGACAACTTATGTATGAC 2004
Qy 724 TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsn 743
Db 2005 TATGCCCGCAAGAACATTCTCTAACTGTGATGAAGCTGCAGAGGAGTTGACCCGA 2064
Qy 744 GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
Db 2065 GGATATGTTGAGCTGAGAAAGCAGCGAAGTTTGTGGCAGTAGCAAAAGGTCTATAACA 2124
Qy 764 AlaThrProArgGlnLeuSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
Db 2125 GCAACTCTTAGGAGATGAAAGTTTGTATGATGAGTGTAGTGAGGCTTGGCTCGGATGCC 2184
Qy 784 PheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
Db 2185 TTCTCGAATGGTTGAAAGCATGACGTGGACGAGCATTTTCGACTTCTCAGAGTTGCA 2244
Qy 804 LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
Db 2245 ATGCAGCAATCAGCAACGATCAGCCACAGGACCATTTGACATGGACCTGATAAATACT 2304
Qy 824 GlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaLeuAlaLeuArgGluLeu 843
Db 2305 GGAGTGTTCAGCAAGTGAACGAATAGGCGGGATACCTTTGGCTGTCATATAGAGACATA 2364
Qy 844 IleAlaAspLysLysLeuSerProGlySerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
Db 2365 GCTCTGGAGAAATGCAAAATTGGAGGATCATCA---ATGGCTATTCGAGTTACTTGAA 2421
Qy 864 AspIleArgSerGlnSer-----SerValAspValSerLeuGlnAspIleLysAsn 880
Db 2422 GAATGAGAGCATGAGGCAACATAAACCAGAAATTCATCTTCATGATGTAGAAAA 2481
Qy 881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspLeuValLysArg 900
Db 2482 GCAGTCGCGACGCTAGCGAGTGAAGGATTTCTGTAGCTGAAGGTGATAGAAATCAAGAGA 2541
Qy 901 Val 901
Db 2542 GTA 2544
```

RESULT 2
CL959068

```
LOCUS CL959068 3168 bp DNA linear GSS 21-SEP-2004
DEFINITION OSIFCC002182 Oryza sativa Express Library Oryza sativa (indica
cultiivar-group) genomic, genomic survey sequence.
ACCESSION CL959068
VERSION CL959068.1 GI:52372893
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
REFERENCE 1 (bases 1 to 3168)
AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,B.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
Location/Qualifiers
1..3168
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
```

Alignment Scores:
Pred. No.: 9,8e-191 Length: 3168
Score: 2337.50 Matches: 492
Percent Similarity: 67.36% Conservative: 123
Best Local Similarity: 53.89% Mismatches: 175
Query Match: 50.72% Indels: 125
DB: 10 Gaps: 14

US-10-768-511-8 (1-901) x CL959068 (1-3168)

Qy 7 LeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGluGlyValSerThrPro 26
Db 746 TTGGTTTGTGAGCGCTGTGTCGCG-----GATGTCGCGCGTCGAGCCCT 793
Qy 27 LeuProGlnValThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArg 46
Db 794 CTCCCGCGGAC-----AACTCTCTCCGCTCAGTCGGGGCGCGC 835
Qy 47 AlaValArgGlnThrProThr-----SerAlaValArgArgGlyArgGluThr 63
Db 836 GCGCGGTGGCGCGCGCGCGGTGGCTCCGCCACCGGTACC-CGTCTCCGCTCCCTCG 894
Qy 64 AspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSer 83
Db 895 GCGGGTTCAGACGCGCGCACCCGCGCCCGCGCACTC----- 933
Qy 84 ProTyrAspAla-----GlyThrProGlyThrProGlyThrProValAla 98
Db 934 CTCTCG 993
Qy 99 ThrProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGly 118
Db 994 TTCCCG-----CCGACCCCGTCCACCCCATGTCC 1023
Qy 119 ThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArg 138
Db 1024 ACCGACGATGTCCGCTCTCTCGAGGCGCGGGA-CGAGGACACCCCGGACCGCG 1082

Qy	14	SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro	33
Db	126	AGCAACCAACGACGACCATGTGTCCTCCCGCATCCACCCCG	164
Qy	34	SerPheAspAlaAlaSerProValAlaGlyArgAlaValArgGlnThrProThr	53
Db	165	-----AGCGCCGACGAGCCGACGCGGA---CGAGTCACCCCAACC	203
Qy	54	SerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArgSerArgSer	73
Db	204	CAGTCCCTTCGAAGTGAAGAGCAGTGTCCACCCAATCGAGACGTAGAGCGGAAGAT	263
Qy	74	ArgSerLeuGly-----	77
Db	264	TCTTCCACCGAGAGCTACTGCCAATGCCACCTCACCAGGA3CCGACCTGCAGAGCCCA	323
Qy	78	-----AenSerValTyrSerSerPro-----	84
Db	324	CCTGCACAGATGCTTGTGTTTCAGGCCCTCCTCAGATGCATCTTTAGTATATTCCTTTG	383
Qy	85	---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla	103
Db	384	GACTTGATGTAGTTCAACATTGACA-----TATGGC	416
Qy	104	ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr	122
Db	417	ACTCCAGCTCGCAGTGGAGGAACCCCAAGAGTGGGGTGAAGGACACACCT---GTA	473
Qy	123	LysGlnArgSerGluLeuGlySerGlnGlyLeuProLeuHisArgArgArgSerGln	142
Db	474	AGGCAGAGCCAGATCTGGCTCAGCAGCAAGAGGTTTCAGATGGATCTGCAGTCT	530
Qy	143	SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer	162
Db	531	-----GATGGCGCAGCAGCA	545
Qy	163	GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGlyThrAsn	182
Db	546	GAACACATCGTACCAGTGAACAGCTCTCTAGGCCAAAAGCTTCTGTGATTTGGGGAACAGAT	605
Qy	183	ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr-----	199
Db	606	GTGAATGTGGCAACATGTAAGAGAAATTTTCAGAGATTCCTTCAGTGTTTTCTGATCTCT	665
Qy	200	-----ArgSerSerAlaHisAspLeuAsnSerIys---TyrIleGlnIle	213
Db	666	CTGGCCAAAGAAGAAATGTTGGCATAGATATTACTCAACCTTTGTATACGTCAACAA	725
Qy	214	IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr	233
Db	726	CTTCGAGAGATTAATATTATACAGGAGAGCCATTTTAAATGTGAATCTCGCAACACATAAAA	785
Qy	234	AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu	253
Db	786	TCATTAGCAAAATCTGTATAGACAGCTCATCTCTACCCACAGGAGGTTATACCAACC	845
Qy	254	LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro-----ThrPheGluLys	271
Db	846	TTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTTAGAACAT	905
Qy	272	HisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPro	291
Db	906	CAGATTCAAGTCACAGACCTTTTAAATGCGTGTGAAGACAAAGAGATATGAGAACTTTGAATCCA	965
Qy	292	SerAspIleAspLysLeuValSerValLysGlyMetValIleAArgCysSerSerIleIle	311
Db	966	GAAGACATTGATCAGCTCATCACCATCAGTGGCATGTCTATCAACAAATCACAGCTGATTT	1025
Qy	312	ProGluLeuLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVal	331
Db	1026	CCGAGAGATGACGAGGCGCTTTTTCGAATGCCAAGCTGTGCGCCACACC-ACC CGGGTG	1082

QY	332	ThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla	351
DB	1083	GAGATAGATCGAGCGAAGATTGCTGAGCCCTGCAGCTTGT-----GTGCACCTGCCACACT	1136
QY	352	ArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArg	371
DB	1137	ACCACAGCATGCGACTGATCCACACCGGATCAATCTTCTGTACAAGCAAAATGATCAAA	1196
QY	372	LeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCys	391
DB	1197	CTTCAAGAGTCTCTGNAGACATGCTGTGGCGACACACCTCACACTATGTCTTTT	1256
QY	392	LeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVal	411
DB	1257	GCCCACAATGACCTTGTGTGACAAGGTTCAACACAGGGGACAGAGTGAACGTCACAGGCATA	1316
QY	412	PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys	431
DB	1317	TATCGACGACTACCAATTTCGAGTTAATCCAAAGAGTGAGCAACGCTGAAGTCTGTCTATAAA	1376
QY	432	ThrTyrIleAspCysValHisValLysValSerAspArgGlyArgLeuGlnThrGluAsp	451
DB	1377	ACCCACATTGATGTCATTCAATATCGGAACCGATGCAAAACGTCTGCAT-----	1427
QY	452	ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer	471
DB	1428	-----GGCCTTGATGAAGAAGCAGAACAG	1451
QY	472	GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle	491
DB	1452	AAACTTTTTCAGAGAAACGCTGTGAATTGCTTAAGGAACCTTTCAGGAGCCAGCATATT	1511
QY	492	TyrAspArgLeuSerArgSerLeuAlaProSerIleTnrcGluLeuGluAspIleLysLys	511
DB	1512	TATGAGCGGCTTGCTCAGCCCTTGGCTCCAGCATTTATGAACATGAAGATATCAAAAG	1571
QY	512	GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle-----ProSerGlyAla	529
DB	1572	GGAACTTACTTCAGCTCTTTGGTGGNAACAAAGGAAGATTTCAGTCACACTGGGAGGGGT	1631
QY	530	SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln	549
DB	1632	AAATTCCGTCGTGAGATCAACATCCTCTCTGTGTGGGAGCCCTGGCACCAAGCAAGTCCCAG	1691
QY	550	LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer	569
DB	1692	CTGCTACAGATATGTATCAACCTGGTGGTCCCAAGGCGCAGTACACGCTGTGAAAGGCTCC	1751
QY	570	SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu	589
DB	1752	AGTGGGTTCGGCTCACCGCTATGTGATGAAGAAGACCTTGAGACAGGACGCTTGTCTCTC	1811
QY	590	GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys	609
DB	1812	CAGACAGTGCCTCGTCTGAGTGCAATGGGATATGCTGCATCGATGAGTTTGACAA	1871
QY	610	MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal	629
DB	1872	ATGAATGAAGCAAGAGTCTGTGCTGATGAGGTATGGAAACAGCAGACTCTGTCTCAT	1931
QY	630	AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn	649
DB	1932	GCAAGGCTGGGATCATCTGTACGTTCAATGCGGACCTCTGTCTCTGGCAGCAGCAAT	1991
QY	650	ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro	669
DB	1992	CCTATTGAGTCTCAGTGGAACTCTAAAAAACAACCACTTGAAAAATATCCAACCTACCGCAC	2051
QY	670	ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn	689
DB	2052	ACATTGTGTCAAGGTTTGATCTCATTTTCTCATGTCAGACCCCTCAGGATGAGGCATAT	2111
QY	690	AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys	709

DIDLITISGMVIRTSQILPEMQEAFQCVCAHTTRVEIDRGRIAEPCSCVCHHTH
 SMALHNSPFSQMKIQSPEDMPAGOTPHITVLFAHNDLVDKVQPGDRVANVTGI
 YRAPVIRPNRVSNNKSVYKTHIDVIHVKRTDAKELHGLDEAEQKLFSEKRVLLKE
 LSRPFDIYERLASALAPISYEHEDIKKGILLQFGTKRDFSGRKFRAEINILLC
 NGPFSKQLQYVNLVPRGQYTGKSSAVGLTAYVKDPETROLVLQVLTGALVLS
 GDCIDKDFKNSTSRVSLHVEVMEQQTLSIAKACIICQLNARTSVLAANPIESOWN
 PKKTIENIQPLHTLLSRFDLITFLMDPQDEAYDRFLAHLHVLVSYQSEEEVEBEFLD
 MAVLKDYTAYAHSTIMPLRSEASQALJEAYVNMKGKSSRGMVSAYPROLESILRLA
 EAHAKVRSNKVEALDVEAKLHREALKQSATDERTGMVDISILITGMSATSRKKE
 ELAENLRKLLISKGTALKYQQLFEDIRGQSDTAITKDMFEALALADDLFTVTG
 KTVRLLL"

polyA_signal

3268..3273

/note="putative"

polyA_site

3285

/note="putative"

ORIGIN

Alignment Scores:

Pred. No.:	8,21e-146	Length:	3235
Score:	1818,50	Matches:	402
Percent Similarity:	59.57%	Conservative:	152
Best Local Similarity:	43.23%	Mismatches:	267
Query Match:	39.46%	Indels:	109
DB:	4	Gaps:	17

US-10-768-511-8 (1-901) x AK089999 (1-3285)

QY	14	SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro	33
DB	110	AGCAACACAGCAGCAGCATGCTGCTCCCGCATCCACCCCG	148
QY	34	SerPheAspAenAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr	53
DB	149	-----AGCGCGCAGCAGCCGCGGA---CGAGTCACCCCAACC	187
QY	54	SerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArgSerArgSer	73
DB	188	CAGTCCCTCGAAGTGAGGAAAGCAGTCGTCACCCCAATCGGACGCTAGAGCGAAGAT	247
QY	74	ArgSerLeuGly-----	77
DB	248	TCTTCCACCGGAGAGCTACTGCCAATGCCACCTCACCAGGAGCGGACCTGCAGAGCCCA	307
QY	78	-----AsnSerValTyrSerSerPro-----	84
DB	308	CCTGCACAGATGCTGTGTTTCCAGCCCTCCTCAGATGCATTTTATAGCTATTCTCTTG	367
QY	85	---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla	103
DB	368	GACTTTGATGTAGTTCACCATTTGACA-----TATGGC	400
QY	104	ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr	122
DB	401	ACTCCAGCTCGCAGTGTGAAGAACCCCAAGAGTGGGGTGACAGGCACACCT---GTA	457
QY	123	LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGln	142
DB	458	AGGCAGAGCCAGATCTGGCTGCACACGAAGAGGTTTGCAGGTGCAGTCTCGAGTCT---	514
QY	143	SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer	162
DB	515	-----GATGGCGCAGCAGCA	529
QY	163	GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTyrGlyThrAsn	182
DB	530	GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAGAGCTGTGATTTGGGGAACAGAT	589
QY	183	ValaenlleProAspValleuArgAlaIleArgArgPheLeuHisAsnTyr-----	199
DB	590	GTGAATGTGGCAACATGTAAGAGAGATTTTTCAGAGATTCCTTCAATGTCTTACTGATCCT	649
QY	200	-----ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnlle	213
DB	1676	CTGCTACAGTATGTGTACAACTGTGTGCCAGAGGCCAGTACACGCTCTGGAAAGGCTCC	1735
DB	650	CTGCCCAAGAAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAA	709
QY	214	IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr	233
DB	710	CTTGAGAGATTAATATTACAGGAGCCATTTTAAATGTGAACCTGCGAACACATAAAA	769
QY	234	AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu	253
DB	770	TCATTTAGCAAAAATCTGTATAGACAGCTCATCTCTACCCACAGGAGGTATATACCAACC	829
QY	254	LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro-----ThrPheGluLys	271
DB	830	TTTGACATGGCTGCAATGAGATCTTCTTTCACCGTTATCTGACTCCATCTTAGAACAT	889
QY	272	HisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPro	291
DB	890	CAGATTCAAGTCAGACCTTTTAAATGCGTTGAAGACAAAGAGATATGAGAAACTTGAATCCA	949
QY	292	SerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIle	311
DB	950	GAAGACATTTGATCAGCTCATCAGTCAGTGGCATGTCATCAGAACATCACAGCTGATT	1009
QY	312	ProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVal	331
DB	1010	CCGAGATGCGAGGAGGCTTTTCCATGCGCAAGTCTGTGCCACACC---ACCCGGGTG	1066
QY	332	ThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla	351
DB	1067	GAGATAGATCGAGGAGAAATGCTGAGCCCTGCGAGTTGT-----GTGCACCTGCCACACT	1120
QY	352	ArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArg	371
DB	1121	ACCCACAGATGCGACATGATCCACACCGATCATCTTCTCTGACAGCAATGATCAAAA	1180
QY	372	LeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCys	391
DB	1181	CTTCAAGAGTCTCTGGAAGACATGCTGCTGGCAGACACCTCACACTATTGCTCTTTT	1240
QY	392	LeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVal	411
DB	1241	GCCACACATGACCTTTGTTGACAAGTTTCAACAGGGGAGAGAGTGAACGTTCAGGCAATA	1300
QY	412	PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys	431
DB	1301	TATCAGCAGTACCAATTCGAGTTAATCCAGAGTGAAGCAAGTGAAGTCTGTCTATATAA	1360
QY	432	ThrTyrIleAspCysValHisValLysSerAspArgGlyArgLeuGlnThrGluAsp	451
DB	1361	ACCCACATTTGATGTCATTCATTATCGGAAAACGGATGCAAAACGCTTGCTGAT-----	1411
QY	452	ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer	471
DB	1412	-----GGCCTTGTGAAGAGCAGAACAG	1435
QY	472	GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle	491
DB	1436	AAACTTTTTCAGAGAAACGTGTGAATGCTTAAAGAACTTCCAGGAGCCAGATATT	1495
QY	492	TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys	511
DB	1496	TATGAGCGCTTGTTCAGCCTTGGCTCCAGCATTTATGAACATGAAGATATCAAAAAG	1555
QY	512	GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysIle-----ProSerGlyAla	529
DB	1556	GGAAATCTTACTTACGCTCTTTGGTGGAAACAAGGAGGATTTTCAGTCACACTGGGAGGGT	1615
QY	530	SerPheArgGlyAspIleAsnValLeuValGlyAspProGlyThrSerLysSerGln	549
DB	1616	AAATTCCTGCTGATGATCAACATCTTCTGTGTGGGACCTTGGCAGCAGCAAGTCCAG	1675
QY	550	LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgLys	569
DB	1676	CTGCTACAGTATGTGTACAACTGTGTGCCAGAGGCCAGTACACGCTCTGGAAAGGCTCC	1735


```

URL:http://fantom.gsc.riken.jp/.
FEATURES
  source
    Location/Qualifiers
      1..3545
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6J"
        /db_xref="PANTOM.DB:9830146H02"
        /db_xref="taxon:10090"
        /clone="9830146H02"
        /sex="male"
        /tissue_type="bone"
        /clone_lib="RIKEN full-length enriched mouse cDNA library"
        /dev_stage="adult"
      misc_feature
        180..2769
          /note="mini chromosome maintenance deficient 4 homolog (S. cerevisiae) (MGDI:103199, GB|NM_008565, evidence: BLASTN, 100%, match=3221)"
      polyA_signal
        3525..3530
          /note="putative"
      polyA_site
        3545
          /note="putative"
    ORIGIN
      Alignment Scores:
      Pred. No.: 5,51e-145 Length: 3545
      Score: 1809.50 Matches: 403
      Percent Similarity: 59.51% Conservative: 151
      Best Local Similarity: 43.29% Mismatches: 267
      Query Match: 39.26% Indels: 11C
      DB: 4 Gaps: 17
      US-10-768-511-8 (1-901) x AK036653 (1-3545)
      QY 14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
      DB 165 AGCAACACGACGACCATGTGCTCCCGCGCATCCACCCCG----- 203
      QY 34 SerPheAspAsnAlaSerProValAlaGlyArgAlaValArgGlnThrProThr 53
      DB 204 -----AGCGCGCGCAGCAGCGCGGGA---CGAGTCAACCCCAACC 242
      QY 54 SerAlaValArgArgGlyArgGluThrAspSerAlaArgArgArgSerArgSer 73
      DB 243 CAGTCCCTCGAAGTGAGGAAAGCAGGTGCTCACCCTCGGAACGAGTACGAGCGG 302
      QY 74 ArgSerLeuGly----- 77
      DB 303 TCTTCCCGGAGAGCTACTGCCAATGCCACCTCACCAGGAGCGGACCTGCAGAGCCCA 362
      QY 78 -----AsnSerValTyrSerSerPro----- 84
      DB 363 CTGTCACAGATGCGCTGTGTTCCAGCCCTCCTCAGATGCAATTTTAGCTATTCCTTTG 422
      QY 85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
      DB 423 GACTTGTAGTTAGTTCACCATGACA-----TATGCG 455
      QY 104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122
      DB 456 ACTCCAGCTCCGAGTGGAGAACCCCAAGATGGGGTGAGAGGACACCT---GTA 512
      QY 123 LysGlnArgSerGluLeuGlySer-GlnGlyLysProLeuHisArgArgArgSerGln 142
      DB 513 AGGCAGAGGCCAGATCTGGCTCAGCCAGAAAGGTTTTCAGAGTGGATCTCAGTCT-- 570
      QY 142 nSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSe 162
      DB 571 -----GATGGCGCAGCAGC 584
      QY 162 rGluSerAlaGluProAspThrLeuGlyGlyTyrAlaTyrValTyrGlyThrAs 182
      DB 585 AGAAGACATCGTACCAAGTGAACAGTCTCTAGGCCCAAAAGCTTGATTTGGGGACAGA 644

```

```

QY 182 nValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr----- 199
DB 645 TGTGAATGTGGCAACATGTAAAGAGAAATTTTCAGAGATTCCTTCAGTGTCTTACTGATCC 704
QY 200 -----ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIle 213
DB 705 TCTGGCCAAAGAAAGAAATGTGGCATAGATATTACTCAACCTTTGTATCATGCAACA 764
QY 213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
DB 765 ACTTGGAGAGATTAATATTACAGGAGAGCCATTTTAAATGTGAATCGGAACACATATAA 824
QY 233 rAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProle 253
DB 825 ATCAATTTAGCAAAAATCTGTATAGACAGCTCATCTCCACACAGGAGTTATACCAAC 884
QY 253 uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro-----ThrPheGlu 271
DB 885 CTTTGACATGGCTGTCAATGAGATCTTCTTGACCGTTATCTGACTCCATCTTAGAACA 944
QY 271 sHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPr 291
DB 945 TCAGATTCAGCTCAGACCTTTTAAATGCTTGAACACAAAGAGATGAGAAACTTGAATCC 1004
QY 291 oSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIle 311
DB 1005 AGAAGACATTCATCAGCTCATCAGTCAGTGGTATGTCATCAGAACATCAGCTGAT 1064
QY 311 eProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVa 331
DB 1065 TCCGAGATGCGAGGAGGCTTTTCCATGCAAGTCTGTGCCCAACCC---ACCCGGGT 1121
QY 331 lThrValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla 351
DB 1122 GGAGATAGATCGAGCGAGAAATGCTGAGCCCTGCAGTGTG-----GTGCACTGCCACAC 1175
QY 351 aArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValAr 371
DB 1176 TACCACAGCATGCGACTGATCCACACCGATCATTTCTCTGCAACAGCAAAATGATCAA 1235
QY 371 gLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCy 391
DB 1236 ACTTCAGAGTCTCCTGAAGACATGCTGCTGGCGAGACACCTCACATATGTCTCTTT 1295
QY 391 sLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVa 411
DB 1296 TGCCCAACATGACCTTGTGACAAAGTTCAACAGGCGACAGAGTGAACGTCACAGGCAT 1355
QY 411 lPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrIly 431
DB 1356 ATATCGAGCAGTACCAATTCGAGTTAATCCAAAGAGTGAACACGCTGAAGTCTGTCTATAA 1415
QY 431 sThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAs 451
DB 1416 AACCCACATTCATGTCATTTCATTCGGAACCGATCGAAACGCTGTCAT----- 1467
QY 451 pProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSe 471
DB 1468 -----GGCCTTGATGAAGAGCAGCAACA 1490
QY 471 rGluAlaAlaAsnGluAlaLysIleGlnLysGluLeuSerLysLeuProGlyIle 491
DB 1491 GAAACTTTTTTCAGAGAAACGTTGAAATTCGTTAAAGAACTTTCCAGGAGCAGATAT 1550
QY 491 eTyrAspArgLeuSerArgSerLeuAlaProSerIleTyrGluLeuGluAspIleIleV 511
DB 1551 TTATAGCGGCTTGCTTCAGGCTTGGCTCCAGCAGCTTTATGAACATGAAGATATCAAAA 1610
QY 511 sGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle-----ProSerGlyAl 529
DB 1611 GGAATCTTACTTCAGCTCTTTGGTGGAAACAAAGGAAGGATTTTCAGTCACACTGGGAGGG 1670

```

QY 529 aSerPheArgGlyAspIleAenValLeuLeuValGlyVasppProGlyThrSerLySerG1 549
Db 1671 TAAATTCGGTGTGAGATCAACATCTCTGTGTGGGACCTCTGGCACAGCAAGTCCCA 1730
QY 549 nLeuLeuGlnThrValHisLeysIleAlaProArgGlyIleThrSerGlyArgGlyse 569
Db 1731 GCTGCTACATGATGTGACCACTGGTCCCGAGAGCCAGTACAGCTCTGGAAAAGGCTC 1790
QY 569 rSerAlaValGlyLeuThrAlaThrValThrLyshAppProGluThrArgGluThrValle 589
Db 1791 CAGTCGGTGGCCCTCACCGCTATGTGATGAAGACCCCTGAGACGACGCTGTGTCT 1850
QY 589 uGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspIly 609
Db 1851 CCGACAGGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1910
QY 609 sMetSerAspAenAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVa 629
Db 1911 AATGAATGAAGACACAGGCTCTGTGCTGATGAGTCTGGAACAGCAGACTCTGTCCAT 1970
QY 629 lAlaLyGlyGlyIleAlaSerLeuAenAlaArgThrSerValLeuAlaCysAlaAs 649
Db 1971 TGCAAAAGGTGGGATCATCTGTGAGTCAATCGGGCACCTCTGTCTGGCAGACGAAA 2030
QY 649 nProSerGlySerArgTyAsnAlaArgLeuSerValIleAspAenIleGlnLeuProPr 669
Db 2031 TCCATTATGATGCTCAGTGGAAATCTTAAACCAACCATTTGAATATATCAACTACCGCA 2090
QY 669 oThrLeuLeuSerArgPheAspLeuIleTyLeuMetLeuAspIlyProAspGluGlnAs 689
Db 2091 CACATTGTGTCAGGTTTGATCTCTATTTCTCTGCTAGACCTCAGGATGAGGCATA 2150
QY 689 nAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyGluAenTyGluValSerIy 709
Db 2151 TGACCGGCTGTAGCTCATCCTGCTTTTCAATGTACTACCAAGGAGGAGCAAGTGA 2210
QY 709 sGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyIleThrTyAlaArgGlnHisVa 729
Db 2211 CGAGGAGTTCCTGGACATGGCTGCTGAAAGACTACATTCATATATGCCATAGTACCAT 2270
QY 729 lHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAenGlyTyTrValGluMetAr 749
Db 2271 CATGCCCCGACTGAGTGAGGAGCCAGCAGGCTCTCATGAGGCTTATGTAACATGAG 2330
QY 749 gGlnLyGlyAsnPheProGlySerSerLyLyValIleThrAlaThrProArgGlnLe 769
Db 2331 GAAG-----ATTGGGAGTAGCGGGGATGGTTCTGCTTACCCTCGACAGCT 2378
QY 769 uGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValG1 789
Db 2379 AGAGTCATTAAATTCGCTTAGCAGAGCCCATGCTAAAGTAAGATTTTCAACAAAGTTGA 2438
QY 789 uLyValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaTh 809
Db 2439 AGCAATTGATGTGGAGAGCAAAAGCCCTCCACCGGAGGCTCTGAAGCAGTCTCAAC 2498
QY 809 rAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerG1 829
Db 2499 TGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2558
QY 829 uArgIleArgArgAlaAenLeuAlaAlaLeuArgGluLeuIleAlaAspIlyIleSe 849
Db 2559 TCGTAACGGAAGAGAGATTAAGTACGAGCAATTTGAGGATATTTGAGGATATTTGAGG 2609
QY 849 rProGlySerSerSerGlyLeuLyThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
Db 2610 TAAGGGTAAACACCCAGCTTAAAGTACCAACAGCTGTTTGGAGATATTTGAGGATATTTGAGG 2669
QY 869 rSerValAspValSerLeuGlnAspIleLyAsnAlaLeuGlySerLeuGlnGlyGluG1 889
Db 2670 TGACACAGCAATTAACCAAGGACATGTTTGAAGAGCCCTCGAGCTTTGGCTGATGATGA 2729
QY 889 yPheLeuThrValHisGlyAspIleValLyS 899

Db 2730 TTCTCTACAGTCTGACTGGGAAGACTGTCCGC 2760
RESULT 6
AK011743 LOCUS
DEFINITION
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610042115 product:mini chromosome
maintenance deficient 4 homolog (S. cerevisiae), full insert
sequence.
AK011743 3316 bp mRNA linear HTC 03-APR-2004
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610042115 product:mini chromosome
maintenance deficient 4 homolog (S. cerevisiae), full insert
sequence.
AK011743
AK011743.1 GI:12848060
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3316)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222.


```
QY 275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294
Db 705 GTCAGACCTTTTAATGCGTTGAAGCAAGAGATATGAGAACTTGAATCCAGAAACATT 764
QY 295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
Db 765 GATCAGCTCATCACCATTGATGCGCATGGTTCATCAGAACATCAGCTGATTCGGAGATG 824
QY 315 LysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValVal 334
Db 825 CAGGAGGCGCTTTTCCAAATGCCAAGTCTGTGCCACACCC---ACCCGGGTGGAGATAGAT 881
QY 335 LysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAla 354
Db 882 CGAGCGAATTTGCTGAGCCCTGCGAGTTGT-----GTGCACTGCCACCTACCCACAGC 935
QY 355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374
Db 936 ATGGCACTGATCCACACCGATCATCTTCTCTGACAGCAAAATGATCAAACTTCAAGAG 995
QY 375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394
Db 996 TCTCCTGAAGACATGCTGCTGGCGAGACACCTCACATATTGTCTTTTGGCCCAAT 1055
QY 395 ThrMetValAspAlaValSerProGlyAspArgIleGluValThrGlyValPheLysAla 414
Db 1056 GACCTTGTGACAGTTTCAACAGGCGGACAGAGTGAACGTGCACAGGCATATATCGAGCA 1115
QY 415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434
Db 1116 GTACCAATTCGATTAATCCAGAGTGCAGCAACGTGAAGTCTGTCTATTAACCCACATT 1175
QY 435 AspCysValHisValLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db 1176 GATGTCATTCATTATCGGAAACGATGCAAAACGCTCTGCAT----- 1217
QY 455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerLysThrSerGluAlaAla 474
Db 1218 -----GGCCTTGATGAAGAACAGACAGAACTTTT 1250
QY 475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db 1251 TCAGAGAAACGTGTGAATGCTTAAGGAACTTTCAGGAACCGATATTTATGAGCGG 1310
QY 495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514
Db 1311 CTTGCTTCAGCCTTGCTCCAGCATTTATGAACATGAAGATATCAAAAAGGGAATCTTA 1370
QY 515 CysGlnLeuPheGlyGlyLysAlaLysLysIle-----ProSerGlyAlaSerPheArg 532
Db 1371 CTTACGCTCTTTGGTGGAAACAGGAAGGATTTCACTCACACTCGGAGGGGTAAATTTCCGT 1430
QY 533 GlyAspIleAsnValLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGln 552
Db 1431 GCTGAGATCAACATCTTCTGTGTGGGACCCCTGGCACCGAGAGTCCAGCTGCTACAG 1490
QY 553 TyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaVal 572
Db 1491 TATGTGTACAACCTGGTCCCAAGAGCCAGATACACTGTCTGGAAGAGGCTCCAGTGGGTC 1550
QY 573 GlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGly 592
Db 1551 GGCCTCACCGCCTATGTGATGAAGACCCCTTGAGACCGAGCAGCTGTCTCTCCAGCAGGT 1610
QY 593 AlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAsp 612
Db 1611 GCCCTGCTGCTGAGTGACAAATGGGATATGCTGCATCGATGATGATTTGACAAATATGAATGA 1670
QY 613 AsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGly 632
Db 1671 AGCAACAAGTCTGTGCTGATGAGTTCATGGAAACAGCAGCTCTGTCCATTGCCAAGGCT 1730
```

```
QY 633 GlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGly 652
Db 1731 GGGATCATCTCTGAGCTCAATGCGCGCACCTCTCTCTCTGGCAGCAGCAATCTTATTGAG 1790
QY 653 SerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeu 672
Db 1791 TCTCAGTGAATCTCTAAAAAACAACCATTTGAAATATCCAACCTACCGCACACATTGTTG 1850
QY 673 SerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArg 692
Db 1851 TCAAGGTTGATCTCATTTCTCTCATGTAGACCTCTCAGGATGAGGCATATGACCGCGGT 1910
QY 693 LeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAla 712
Db 1911 CTAGCTCATCACCTGGTTTCTTCAATGTGTACCAAGTGAAGGAGCAAGTGGAGGAGGTTC 1970
QY 713 LeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrValArgGlnHisValHisProThr 732
Db 1971 CTGGACATGGCGCGTCTGAAAGACTACATTGTCATATGCCATAGTACCATCATGCCCGA 2030
QY 733 LeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGly 752
Db 2031 CTGAGTGAGGAGCGCAGCCAGGCTCTCATTCAGGCTTATGTAAACATCAGGAAG----- 2084
QY 753 AsnPheProGlySerSerLysValIleThrAlaThrProArgGlnLeuGluSerMet 772
Db 2085 -----ATTGGAGTAGCCGGGGATGGTTTCTGCTTACCTTCGACACTAGAGTCATTA 2138
QY 773 IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAsp 792
Db 2139 ATTGCTTAGCAGAAGCCCATGCTTAAAGTAAGATTTTCAAAACAAAGTTGAAGCAATTGAT 2198
QY 793 AlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAla 812
Db 2199 GTGGAAGAGGCAAAACGCTCCACCGGAGGCTCTGAAGCAGTCTGCAACTGACCTCGT 2258
QY 813 ThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArg 832
Db 2259 ACTGGCATTTGGGATATTTCTATTCTTACTACAGAATGAGTGGCCACTTCTCGTAAACGG 2318
QY 833 ArgAlaAsnLeuLeuAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySer 852
Db 2319 AAAGAGAAATTAGCTGAAGCATTCAGAAAACCTATT-----TTATCTAAGGGTAA 2369
QY 853 SerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAsp 872
Db 2370 ACACGAGCCTTAAAGTACCAACAGCTGTTTTCAGGATATTCGGGGACAGCTCTGACACAGCA 2429
QY 873 ValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThr 892
Db 2430 ATTACCAAGGACATGTTTGAAGAGCCCTCGCAGCTTTGGCTGTGATGATTTCTTAACA 2489
QY 893 ValHisGlyAspIleValLys 899
Db 2490 GTGACTGGGAAGACTGTCCGC 2510
RESULT 9
AY408034 2357 bp DNA linear GSS 12-DEC-2003
LOCUS Pan troglodytes HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408034
VERSION AY408034.1 GI:39764005
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE 1 (bases 1 to 2357)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
```

TITLE Adams,M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302

REFERENCE 2 (bases 1 to 2357)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.

TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering
 them based on alignment.

FEATURES Location/Qualifiers

source
 1..2357
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 gene
 <1..>2357
 /locus_tag="HCM3089"

ORIGIN

Alignment Scores:

Pred. No.: 2,14e-132 Length: 2357
 Score: 1661.00 Matches: 349
 Percent Similarity: 64.01% Conservative: 133
 Best Local Similarity: 46.35% Mismatches: 209
 Query Match: 36.04% Indels: 62
 DB: 10 Gaps: 12

US-10-768-511-8 (1-901) x AY408034 (1-2357)

QY 85 TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThr 104
 DB 15 TTTGATGTAGTTTACCCTGACA-----TACGGCACT 47
 QY 105 ProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyrLys 123
 DB 48 CCCAGCTCTCGGTAGAGGGAACCCCAAGAGTGGTGTAGGGGCACACCT---GTGAGA 104
 QY 124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGlnSer 143
 DB 105 CAGAGCGCTGACCTGGCTCTGCACAGAGGGCGCTCAAGTGGATCTGCAGTCT----- 158
 QY 144 ArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGlu 163
 DB 159 -----GACGGGCGACGACAGAA 176
 QY 164 SerAlaGluProAspAspThrLeuGlyGlyGlyTyrAlaTyrValTrpGlyThrAsnVal 183
 DB 177 GATATAGTGGCAAGCGAGCAGCTCTAGGCCAAAACCTTGTGATCTGGGGACAGATGTA 236
 QY 184 AsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr----- 199
 DB 237 AATGTGGCAGCAGTCAAGAAACCTTTTCAGAGATTCTTCAGCGTTTATTACCCCTCTG 296
 QY 200 -----ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIle 214
 DB 297 GCTAAGAGAGAGAAATGTTGGCATAGATATTACTGAACCTCTATCATGACAGCACTT 356
 QY 215 GluGluThrValIleuArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234
 DB 357 GGGGAGATTAACTTATTGGTGAGCCATTTTAAATGTGAACCTGTGAACACATCAATCA 416
 QY 235 HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu 254
 DB 417 TTTGACAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAAGTATTCCAACTTTT 476
 QY 255 AspThrGluCysGlnGluValAlaThrSerLeuLeuPro-----ThrPheGluHis 272
 DB 477 GACATGGCTGTCAATGAAATCTTCTTTGACCGGTTACCTCTGACTCAATCTTAGAACATCAG 536

QY 273 IleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer 292
 DB 537 ATTCAAGTAAGACCACTTCAACGCATTGAAGACTAAGAAATATGAGAAACCTGAATCCAGAA 596
 QY 293 AspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIlePro 312
 DB 597 GACATTGACCAAGCTCATCACCATTGATGGCATGGTGTATCAGACATCCAGCATCCGATGATCCC 656
 QY 313 GluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProLeuValThr 332
 DB 657 GAGATGCAAGGAGGCTTCTTCAGTGCACAGTGTGCCCACACG---ACCCGGGTGGAG 713
 QY 333 ValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArg 352
 DB 714 ATGGACCGCGCGCGCATTTGCAGAGCCAGTGTGTGGGGCGC-----TGCCACACCACC 767
 QY 353 AsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeu 372
 DB 768 CACAGCATGGCACTCATCCACACCGCTCCCTCTTCTGTGACAGCATGATCAAGCTT 827
 QY 373 GlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeu 392
 DB 828 CAGGAGTCTCCGGAAGACATGCTCGAGGGCAGACCCACACACAGTTATCTCTGTTGCT 887
 QY 393 TyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPhe 412
 DB 888 CACAATGATCTCGTTGACAAGGTCCAGCTCGGGACAGAGTGAATGTTACAGGCATCTAT 947
 QY 413 LysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThr 432
 DB 948 CGAGCTGTGCTTATTCGAGTCAATCCAAAGAGTGAATTAATGAAGTCTGTCTTACAAAACC 1007
 QY 433 TyrIleAspCysValHisValLysSerAspArgGlyArgLeuGlnThrGluAspPro 452
 DB 1008 CACATTGATGATCATTTATTCGGAACCGATGCAAAACGCTCAT----- 1055
 QY 453 MetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGlu 472
 DB 1056 -----GCGCTGTATGAAGACAGACAGAA 1082
 QY 473 AlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyr 492
 DB 1083 CTTTTCAGAGAAACGTTGGAATGCTTAAAGAACTTTCCAGGAAACACAGACATTTAT 1142
 QY 493 AspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysGly 512
 DB 1143 CAGAGGCTTGTCTTACGCTTGGCTCCAAAGCATTTATGAACATGAAGATATTAAGAAGGA 1202
 QY 513 LeuLeuCysGlnLeuPheGlyLysAlaLysLysIle-----ProSerGlyAlaSer 530
 DB 1203 ATTTTGTCTTACGCTCTTTGGCGGAGACAGGAAGATTTTAGTCACTCGAAGGGGCAAA 1262
 QY 531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550
 DB 1263 TTTTCGGCTGAGATCAACATCTTGTGTGTGGCGACCTGGTACCAGCAAGTCCCAGCTG 1322
 QY 551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570
 DB 1323 CTCAGTATGTATCAACCTCGTCCCGGGGCCAGTACACGCTGTGGGAAGGGCTCCAGT 1382
 QY 571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
 DB 1383 GCAGTTGGCTCCTCCTCGCTACGTAATGAAGAACCCCTGAACAGGCGAGCTGTGCTCGCAG 1442
 QY 591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysAlaIleAspGluPheAspLysMet 610
 DB 1443 CAGGTTGCTCTTGTCTGATGACAAACGCACTGCTGTATTCATGATGATTCGACAGATG 1502
 QY 611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630
 DB 1503 AATGAAGTACAAGATCGGTATTGATGAAGTCATGAAGACAGACAGACTCTGTCTCATTCGA 1562

```

QY 631 LysGlyGlyIleAlaSerLeuAsnAlaArgThrSerVaLeuAlaCysAlaAsnPro 650
Db 1563 AAGGCTGGATCATCTGCAGCTCAATGGCGACCTCTGTCTCGCAGCAGCAATCCC 1622
QY 651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProThr 670
Db 1623 ATTGAGTCTCAGTGGATCCTTAAAAAACACCACTTGAACATCCAGCTGCCTCATACT 1682
QY 671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
Db 1683 TTATTATCAAGGCTTTGATTTGATCTTCTCATGCTGGACCTTCAGGACGAGCCCTATGAC 1742
QY 691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsrTyrGluValSerLysGln 710
Db 1743 AGGCGCTGCGCTCACCACTGGCTGCAGCTGTACTACAGAGCGAGGAGGAGGAGGAG 1802
QY 711 AspAlaLeuAspLeuGlnThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
Db 1803 GAGCTCTGGACATGGCGGTGCTAAAGGACTACATTTGCTAGGCACAGACCATCATG 1862
QY 731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
Db 1863 CCGCGCTAAGTGAGGAGCCAGCCAGCGCTCTCATCGAGGCTTATGTAGACATGAGGAAG 1922
QY 751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
Db 1923 -----ATTGGCAGTAGCCGGGGAATGGTTCTGCATATCCCTCGACAGCTAGAG 1970
QY 771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
Db 1971 TCATTAAATCCGTTAGCAGACCCATGCTAAAGTAAAGATTGCTTAAACAAGTTGAGCC 2030
QY 791 ValAspAlaAlaGluAlaValArgLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db 2031 ATTGATGTGAAGAGAGCCAAACCGCTCCANNNNNNNGCTCTGAAGCAGTCTGCAACTGAT 2090
QY 811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
Db 2091 CCCGCGACTGGCATCGTGACATATCTATTCTTACTACG 2129

RESULT 10
LOCUS CR598188 1607 bp mRNA linear HTC 21-JUL-2004
DEFINITION Full-length cDNA clone CS0D012Y118 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).
ACCESSION CR598188
VERSION CR598188.1 GI:50478995
KEYWORDS HTC; CNSLT CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1607)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1607)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1607

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="CS0D012Y118"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.: 2,39e-104 Length: 1607
Score: 1334.00 Matches: 272
Percent Similarity: 70.74% Conservative: 93
Best Local Similarity: 52.71% Mismatches: 127
Query Match: 28.94% Indels: 24
DB: 4 Gaps: 4

US-10-768-511-8 (1-901) x CR598188 (1-1607)
QY 386 HisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArg 405
Db 3 CACACAGTTATCTCTGTTTGTCTCACAAATGATCTCGTTGACAAGGTCCAGCTGGGACAGA 62
QY 406 IleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThr 425
Db 63 GTGAATGTTACAGGCATCTATCGAGCTGTGCTGCTATTTCGAGTCAATCCAAAGAGTGAGTAAT 122
QY 426 LeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValHisLysLysSerAspArgGly 445
Db 123 GTGAAGTCTGCTACAAAACCCACATTGATGTCATTTCATTATCGGAAAACGGATCGAAA 182
QY 446 ArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyr 465
Db 183 CGTCTGCAT-----GGCCTT 197
QY 466 HisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeu 485
Db 198 GATGAAGAGCAGAACAGAAACTTTTTCAGAGAAACGTGGGAATTCCTTAAGGAACAT 257
QY 486 SerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGlu 505
Db 258 TCCAGGAACCCAGACATTATGAGAGGCTTCTTCAGCCTTGGCTCCCAAGCATTTATGAA 317
QY 506 LeuGluAspIleLysLysGlyLeuLeuCysGlnPheGlyGlyLysAlaLysLysIle 525
Db 318 CATGAAGATATAAAGAGGGAATTTTGCTTCAGCTCTTTGGCGGACAAAGGAAGATTTT 377
QY 526 -----ProSerGlyAlaSerPheArgGlyAspIleAsnValLeuValGlyAspPro 543
Db 378 AGTCACACTGGAAAGGGGCAAAATTTTCGGGCTGAGATCAACATCTGCTGTGGCGACCT 437
QY 544 GlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyr 563
Db 438 GGTACCAGCAAGTCCCAGCTGCTGACGTACGTGTACAACTTCGTCGCCAGGGCCAGTAC 497
QY 564 ThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGlu 583
Db 498 ACGTCTGGAAAGGGCTCCAGTGCAAGTGGCTCACTGCGTACGTAAATGAAAGACCCCTGAG 557
QY 584 ThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCys 603
Db 558 ACAAGGCGAGCTGGTCCGACAGAGGTGCTCTTGTCTGAGTGACAAACGGCATCTGCTGT 617
QY 604 IleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGlu 623
Db 618 ATCGATGAGTTCGACAGATGAATGAAGTACAGATCGGTATTCGATGAGTCATGGA 677
QY 624 GlnGlnThrValSerValAlaLysGlyGlyIleAlaSerLeuAsnAlaArgThrSer 643
Db 678 CAGCAGACTCTGTCCATTGCAAGAGCTGGATCATCTGTCAGCTCAATCGCGCACCTCT 737
QY 644 ValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAsp 663

```

```

Db      738  GTCTGGCAGCAGCAAAATCCCATTCAGTCTCAGTGAATCTCTAAATAAACCAACCACTTCAA 797
Qy      664  AaAIIeGlnLeuProThrLeuLeuSerArgPheAspLeuIleThrLeuMetLeuAsp 683
Db      798  AATATCCAGCTGCCATCTTATATCAAGGTTTGATTTGATCTTCTCTGTCTGGAC 857
Qy      684  LysProAspGluGlnAsnAspArgLeuAlaArgHisLeuValAlaLeuHisThrGlu 703
Db      858  CCTCAGGACGAGCCATGACAGGGCTCTGGCTCACCACCTGGTCGCACTGTACTACCAG 917
Qy      704  AsnThrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaThrLeu 723
Db      918  AGCAGGAGCAGCAGCAGGAGGAGCTCTCTGGACATGGCGGTGTCTAAAGCACTACATGCC 977
Qy      724  TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuLeuAsn 743
Db      978  TAGCGGCACAGCACCATCATGCGCGCTAAGTGAGGAAGCCAGCGGCTCTCATCGAG 1037
Qy      744  GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
Db      1038  GCTTATGTAGATGAGGAGGAG-----ATTGGCAGTAGCGGGGAATGGTTCT 1085
Qy      764  AlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
Db      1086  GCATACCCCTCGACAGCTAGAGTCATTAAATCCGCTTAGCAGAGGCCCATGCTAAAGTAAGA 1145
Qy      784  PheSerGluValValGlnLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
Db      1146  TTGTCTTAACAAGATTGAAGCCATTGATGTGGAAGAGGCCAAGCCCTCCATCGGGAAGCT 1205
Qy      804  LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThr 823
Db      1206  CTGAAGCATGCTGCAACTGATCCCGAGTGGCATCTGTGGACATATCTATTCTTACTAGC 1265
Qy      824  GlyValSerAlaSerGluArgIleArgAlaAsnLeuLeuAlaAlaLeuArgGluLeu 843
Db      1266  GGGATGAGTGCCACTCTCGTAAACGGAAGAGAGTAATAGCTGAAGCATTTGAAAAGCTT 1325
Qy      844  IleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
Db      1326  ATT-----TTATCTAAGGGCAAAACACCAGCTCTAAATAATACCAGCACTTTTGTAA 1376
Qy      864  AspIleArgSerGlnSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGly 883
Db      1377  GATATTCTGGGGCAATCTGACATGCAATTAATAAGATATGTTTGAAGAGCACTGCGT 1436
Qy      884  SerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLys 899
Db      1437  GCCTGGCAGATGATGATTTCTGACAGTGAAGAGCCGTGGCGC 1484

```

```

RESULT 11
LOCUS   DQ057540
DEFINITION 3389 bp mRNA linear HTC 08-AUG-2005
ACCESSION DQ057540
VERSION   DQ057540.1 GI:68301743
KEYWORDS  HTC.
SOURCE    Antonospora locustae
ORGANISM  Antonospora locustae
REFERENCE 1 (bases 1 to 3389)
AUTHORS  Williams,B.A., Slamovits,C.H., Patron,N.J., Fast,N.M. and Keeling,P.J.
TITLE     A high frequency of overlapping gene expression in compacted eukaryotic genomes
JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 102 (31), 10936-10941 (2005)
PUBMED   16037215
REFERENCE 2 (bases 1 to 3389)
AUTHORS  Williams,B.A.P., Slamovits,C.H., Patron,N.J., Fast,N.M. and Keeling,P.J.
TITLE     Direct Submission
JOURNAL  Submitted (09-MAY-2005) Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada

```

FEATURES

```

source
1..3389
/organism="Antonospora locustae"
/mol_type="mRNA"
/db_xref="taxon:278021"
/clone="2f18"

ORIGIN
Alignment Scores:
Pred. No.: 6,18e-101 Length: 3389
Score: 1300.00 Matches: 299
Percent Similarity: 59.94% Conservative: 148
Best Local Similarity: 40.08% Mismatches: 247
Query Match: 28.21% Indels: 52
Gaps: 4
17
US-10-768-511-8 (1-901) x DQ057540 (1-3389)
Qy 142 GlnSerArgGluProGlyHisArgSerProSerArgGluPro--SerAlaAspGlyArgP 161
Db 1025 CAATCAATTAACCCATGGATCGTCGAGATCTCGTCAGCTTTTCTTTGATGAAGCAA 1084
Qy 161 roSerGluSerAlaGluProAspThrLeuGlyGlyGluTyrAla----- 176
Db 1085 GCCAGGACATAGCGCAACCGCTGAATACA---CCTACAGAGTACAGCTCCGATACGTCT 1141
Qy 177 -----Ty-ValTyrGlyThrAsnValAsnIleP 186
Db 1142 CGCAGCCTCTAATAGAGACAGAAAGATCAAGTGAATCTGGGGCAGCACCATCAAGTCG 1201
Qy 186 roAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAlaHisAspL 206
Db 1202 AGGAGTCATCGGAATGTTCAAGGAATTTGTGAGGAGCTTT-----GTGGCACACGACA 1255
Qy 206 euAsnSerLysTyrIleGlnIleLeuGluThrValGluArgGluGluAspThrLeuA 226
Db 1256 CACAC---GTTTACATGGAAGAAGATTGCGGCTCTGATTGATACACACAGTCAAGTACTCTTC 1312
Qy 226 snIleAspMetSerAspIleTyrAspHisAspProAsp---LeuTyrAlaAlaValAla 245
Db 1313 TGATGAATTTGCAAG-----CACCTGCCGAGATGCTGTATTAAGCAACTCGAAT 1360
Qy 245 rgyTyrProLeuAspIleLeuProLeuAspThrGluCysGlnGluValAlaThrSerL 265
Db 1361 CGTATCCGACAGAGATATACCCATTTTGTAGAGACTCTCTGAGCGAGGTGGTTTCAGAGC 1420
Qy 265 euLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValH 285
Db 1421 TTGGCGCCAGAGTTCAACTGTGTAGTGAATCAGGTTTATTAATATCGGCGCAACTGT 1480
Qy 285 leMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValI 305
Db 1481 TCATCAGAACATATCCAGCGCGCACATCGACAGAAATTTGTGCGGTGCGAGGATCGTGA 1540
Qy 305 leArgCysSerSerIleLeuProGluIleLysGlyAlaPhePheLysCysLeuValCysG 325
Db 1541 TAAGAGTGCACAAACGGTATGCTTAAATCTTAAGCATTTTATACATGACATCGATGTG 1600
Qy 325 lyHisSerProProLeuValThrValLysGlyArgValGluGluProThrArgCysG 345
Db 1601 GA---AAGGAGTACTTGTGCAAAACATACGTGGCGGTGAATTAACGAGCCCATCTGCG- 1656
Qy 345 luLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheA 365
Db 1657 -----GACTGCGGTGCAAGGTTTGCATACCAACTTGTCTCCAACAAAGATCTTACA 1708
Qy 365 laAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrP 385
Db 1709 TAGATAAACAGATCATTCAGTCCAGAGCTTCTCAGAAATATACAGGACGGAATACGCG 1768
Qy 385 roHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspA 405
Db 1769 CAATGCCAATAACAGTGTTCGCCAAGACGATTTAGTACACATGTGTATCTGGCGACA 1828

```



```
Db 61 GTTTGGAGAGTGGAGCCCTGTTCTGAGTGACAGAGGCATCTGTTGATTGATGAATTT 120
Qy 608 AspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVal 627
Db 121 GATAAAATGCTGTAATGCAAGAGCAGTCGTCGATGAGGTGATGGAACAACAGACTGTT 180
Qy 628 SerValAlaLysGlyGlyLeuAlaSerLeuAsnAlaAArgThrSerValLeuAlaCys 647
Db 181 TCAATGACAAAGCTGGTATATGTCATCTCTTAATGCAAGGACTTCAGTCTTGGCTGT 240
Qy 648 AlaAsnProSerGlySerArgTyrAsnAlaAArgLeuSerValIleAspAsnIleGlnLeu 667
Db 241 GCAAAATCCTAGTGGTTCGGCGCTATAATCGCGCTCTGCTGTGATTGACAAACATACACCTG 300
Qy 668 ProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGlu 687
Db 301 CTCCTCAACCTTCTGTCAGATTTGATTAATCTACTTGAATCTCGACAGGCTGACGAG 360
Qy 688 GlnAsnAspArgArgLeuAlaAArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVal 707
Db 361 CAGACAGATAGCGGTTAGCTAAGCATATGTTCTTACATTTTGGAGCACAGAGATT 420
Qy 708 SerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaAArgGln 727
Db 421 GCTGAGCATGATGTCATAGATCTTGCTACATTAACAGCATATGTGAGCTATGCTGCGCAG 480
Qy 728 HisValHisProThrLeuSerAspGluAlaAlaGluAspLeuLeuAsnGlyTyrValGlu 747
Db 481 AATATTTCAACCAAAATTTATCTGATGAAGCTGCTGAAGAGCTAACCGAGGATGTTGAA 540
Qy 748 MetArgGlnLysGlyAsnPheProGlySerSerLysValIleThrAlaThrProArg 767
Db 541 ATGAGGAAGAAGGAAATTTCCAGGCAGCAGTAAAGAGTTCATCACAGCACCAAGG 600
Qy 768 GlnLeuGluSerMetIleArgLysSerGluAlaLeuAlaAArgMetArgPheSerGluVal 787
Db 601 CAAATTTGAAAGTCTGATAGCCTTAGTGAAGCCTTGGCTCGTATTTCGATTCAGAAACA 660
Qy 788 ValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSer 807
Db 661 GTTGAATCGAGATGTAATCAGGCGATTTCCGCGCTTTTGGAGTGTCTATGACGACATCA 720
Qy 808 AlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAla 827
Db 721 GCACCTGATCACTCCACAGAACTATTGATGATGATCTTATCACCACTGAGTTTCTGCA 780
Qy 828 SerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuAArgGluLeuIleAlaAspLys 847
Db 781 AGTGAAGGATGAGGAGGAGAGTGTGTTATCAGCAACAGCAACGTAATAATGAGATAAG 840
Qy 848 IleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
Db 841 CTACAACTTTGGGGGACCATCA-----ATCGGATGCTAGAG 876
```

RESULT 15

CK210237

LOCUS

DEFINITION

FGAS022036 Triticum aestivum FGAS: Library 5 GATE 7 Triticum

aestivum cDNA, mRNA sequence.

CK210237

CK210237.1 GI:39572627

EST.

Triticum aestivum (bread wheat)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

1 (bases 1 to 1156)

Allard, F., Crosby, W.L., Danyluk, J., Rudes, F., Frick, M., Gaudet, D.,

Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,

Linke, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,

Penniket, C., Roach, J.L. and Sarhan, F.

TITLE
JOURNAL
COMMENT

Functional Genomics of Abiotic Stress In Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.ests@usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [16,727].
Plate: L5B022 row: D column: 10.

FEATURES
source

1..1156
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
/note="Vector: pCMV.SPORT6; Crown and developmental stages
of spike formation in wheat cultivar Norstar. 4 mRNA
populations were combined before constructing the library.
The first mRNA population is from 1cm crown sections after
30 days of cold acclimation. The second is from 1cm crown
sections after 11 days of deacclimation (before
deacclimation plants were fully vernalized for 49 days).
The third is from different developmental stages of spike
formation (5 to 50mm) that still have not emerged from the
leaf (dissection required). The last is from different
developmental stages of spike and seed formation after
having emerged from the leaf (visible). First strand
synthesis in this library was done in the presence of
methylated dCTP thereby protecting from internal cleavage
with NotI."

ORIGIN

Alignment Scores:
Pred. No.: 1156
Score: 1127.00
Percent Similarity: 84.98%
Best Local Similarity: 74.12%
Query Match: 24.45%
DB: 2
Gaps: 0

US-10-768-511-8 (1-901) x CK210237 (1-1156)

Qy 536 AsnValLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db 35 CATATTTGCTTGTGGTGTCTCGAAGCAGCAATCCATCTCTCCAGTACATGCAT 94
Qy 556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db 95 AAATCTCTCTCTCGTGGTATTACACAGTGAAGAGGAGTTCAGCAGTTGGCTTACT 154
Qy 576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db 155 GCTTATGTTGCTAAGGACCTGAACTGGTGAACATGTTCTTGAAGAGTGAGCATTGTT 214
Qy 596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db 215 TTGAGTGACAAAGGTGTTGCTGTATTGATGAGTTGATAGATGCTCTGATATGCCGA 274
Qy 616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIle 635
Db 275 AGCATGCTCATGAGTGTGATGAGCAGCAGACTGTATCTTCCTTGAAGGCTGGAATATT 334
Qy 636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db 335 GCATCTCTTGAATGCTAGGACATCTGTACTAGCATGTGCGCAATCCATCTCGAATCACGTTAC 394

